

**ANÁLISIS BIG DATA DE LOS PATRONES DE MOVILIDAD A PARTIR DE
REGISTROS DE LLAMADA**

CRISTIAN STIVEN VARGAS ROJO

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE INGENIERÍAS
INGENIERIA EN SISTEMAS Y COMPUTACIÓN
PEREIRA
2017**

**ANÁLISIS BIG DATA DE LOS PATRONES DE MOVILIDAD A PARTIR DE
REGISTROS DE LLAMADA**

CRISTIAN STIVEN VARGAS ROJO

**Trabajo de grado para optar el título de Ingeniero de Sistemas y
Computación**

**Asesora
ANA MARÍA LÓPEZ ECHEVERRY
Docente**

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE INGENIERÍAS
INGENIERIA EN SISTEMAS Y COMPUTACIÓN
PEREIRA
2017**

TABLA DE CONTENIDO

| | Pág. |
|--|------|
| INTRODUCCIÓN | 8 |
| 1. GENERALIDADES | 9 |
| 1.1. TÍTULO | 9 |
| 1.2. FORMULACIÓN DEL PROBLEMA | 9 |
| 1.3. JUSTIFICACIÓN | 9 |
| 1.4. OBJETIVOS | 10 |
| 1.4.1. OBJETIVO GENERAL | 10 |
| 1.4.2. OBJETIVOS ESPECIFICOS | 10 |
| 1.5. MARCO REFERENCIAL | 11 |
| 1.5.1. MARCO TEORICO | 11 |
| 1.6. HIPOTESIS | 19 |
| 1.7. TIPO DE INVESTIGACIÓN | 19 |
| 1.8. METODOLOGIA | 19 |
| 1.8.1. UBICACIÓN DEL EXPERIMENTO | 19 |
| 1.8.2. DISEÑO EXPERIMENTAL | 19 |
| 1.8.3. SELECCIÓN DE LA MUESTRA/ DESCRIPCIÓN DE UNIDADES EXPERIMENTALES | 20 |
| 1.8.4. RECOLECCIÓN DE DATOS | 20 |
| 1.8.5. PROCEDIMIENTOS PARA SU EVALUACIÓN | 20 |
| 2. ESTUDIO DEL ESTADO DE ARTE DE LOS ALGORITMOS EXISTENTES PARA EXTRAER INFORMACIÓN DE LOS REGISTROS DE LLAMADA EN UNA RED CELULAR. | 22 |
| 2.1. ALGORITMOS | 22 |
| 2.1.1. MARCO PARA APOYAR EL CONOCIMIENTO DEL CONTEXTO DE MOVILIDAD EN REDES CELULARES [6] | 23 |
| 2.1.2. DISEÑO Y OPTIMIZACIÓN DE RUTAS Y FRECUENCIAS EN EL TRANSPORTE COLECTIVO URBANO, MODELOS Y ALGORITMOS. [7] ... | 25 |

| | | |
|--------|--|----|
| 2.1.3. | LA PREDICCIÓN DE LA MOVILIDAD DE LOS SUSCRIPTORES DE TELÉFONOS CELULARES USANDO EL ALGORITMO MEJORADO DE LA CADENA DE MARKOV [8] | 30 |
| 3. | CARACTERIZACIÓN DE LAS VARIABLES | 38 |
| 3.1. | RUTA PARA LA CARACTERIZACIÓN DE LAS VARIABLES..... | 38 |
| 3.1.1. | IDENTIFICACIÓN DE LAS VARIABLES A INVOLUCRAR EN EL ANÁLISIS..... | 38 |
| 3.1.2. | DEFINICIÓN DE MECANISMOS DE CARACTERIZACIÓN DE LAS VARIABLES..... | 39 |
| 3.1.3. | CARACTERIZACIÓN DE CADA VARIABLE | 41 |
| 3.1.4. | VARIABLES PARA MEJORAR EL ANÁLISIS DE MOVILIDAD..... | 43 |
| 4. | ANÁLISIS DEL PROCESO DE MOVILIDAD | 46 |
| 4.1. | ESTRUCTURA DEL ANÁLISIS..... | 46 |
| 4.2. | ESCENARIO BASE PARA EL ANÁLISIS DE MOVILIDAD..... | 49 |
| 4.2.1. | CLASIFICACIÓN DE LA MOVILIDAD | 49 |
| 4.2.2. | ANÁLISIS DE MOVILIDAD EN BASE A LAS CADENAS DE MÁRKOV 50 | |
| 4.3. | ANÁLISIS DE MOVILIDAD EN BASE A LA MEJORA DE LAS VARIABLES CARACTERIZADAS | 59 |
| 4.3.1. | EVALUACIÓN DE LAS MEDIDAS PROPUESTAS DE MOVILIDAD. 67 | |
| 4.3.2. | DEFINICIONES MATEMÁTICAS DE LAS MEDIDAS DE MOVILIDAD 68 | |
| 4.3.3. | RESULTADOS DE LA APLICACIÓN DE LAS MEDIDAS EN UN ANÁLISIS PREVIO [10] | 72 |
| 4.4. | PROCESO DE CURACIÓN DE DATOS PARA LA VISUALIZACIÓN [13]84 | |
| 5. | CONCLUSIONES | 90 |
| | BIBLIOGRAFIA | 94 |

LISTA DE FIGURAS

| | Pág. |
|--|------|
| Figura 1. Funcionamiento de un sistema celular | 11 |
| Figura 2. Representación de una celda | 13 |
| Figura 3. Localización de antenas rural y urbana..... | 15 |
| Figura 4. Funcionamiento del algoritmo MapReduce | 18 |
| Figura 5. Motivación | 24 |
| Figura 6. Visión general de la arquitectura del sistema..... | 31 |
| Figura 7. Superposición de intervalos en GTPA..... | 36 |
| Figura 8. Proceso de la caracterización | 38 |
| Figura 9. Descomposición de las variables | 40 |
| Figura 10. Porcentaje según la utilización de las variables | 43 |
| Figura 11. Descripción del análisis de movilidad..... | 47 |
| Figura 12. Mapa de trayectoria..... | 49 |
| Figura 13. Mapa de la trayectoria con posibles lugares a dirigirse..... | 51 |
| Figura 14. Predicción de la movilidad..... | 54 |
| Figura 15. Descomposición funcional..... | 55 |
| Figura 16 Arquitectura General | 57 |
| Figura 17. Interfaz en infraestructura vehicular | 58 |
| Figura 18. Interfaz en las redes en movimiento..... | 59 |
| Figura 19. Mapa de la trayectoria de avenida 30 de agosto, referenciado con cuadrículas..... | 60 |
| Figura 20. Correlaciones estimadas para cada uno de los 44 meses de las CDR entre la medida existente del número de movilidad de las torres utilizadas (NTU) y las seis medidas de los grupos A (azul), B (tonos de verde) y C (tonos de rojo). .. | 74 |
| Figura 21. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la medida existente de distancia de movilidad recorrida (DT-SL) y las seis medidas de los grupos A (azul), B (tonos de verde) y C (tonos de rojo)..... | 75 |
| Figura 22. Correlaciones estimadas para cada uno de los 44 meses de las CDR entre la medida existente de distancia máxima de movilidad recorrida (MDT) y las seis medidas de los grupos A (azul), B (tonos de verde) y C (tonos de rojo)..... | 76 |
| Figura 23. Correlaciones estimadas para cada uno de los 44 meses de CDR entre la medida existente del radio de giro de movilidad (RoG) y las seis medidas de los grupos A (azul), B (tonos de verde) y C (tonos de rojo)..... | 77 |
| Figura 24. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la nueva medida del número de movilidad de viajes (NT) que define el grupo A y las ve medidas de los grupos B (tonos de verde) y C (tonos de rojo). | 78 |

| | |
|--|----|
| Figura 25. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la nueva medida de las celdas de la cuadrícula de movilidad visitadas (GCV-R) del grupo B y la otra medida de movilidad del grupo B (SV-R, verde) Cuatro medidas de los grupos A (azul) y C (tonos de rojo)..... | 79 |
| Figura 26. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la nueva medida de los sitios de movilidad visitados (SV-R) del grupo B y la otra medida de movilidad del grupo B (GCV-R, verde), así como los cuatro Medidas de los grupos A (azul) y C (tonos de rojo)..... | 80 |
| Figura 27. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la nueva medida de distancia de movilidad recorrida (DT-R) del grupo C y las otras medidas de movilidad del grupo C (TT-R y GCT-R, tonos de rojo), Así como las tres medidas de los grupos A (azul) y B (tonos de verde)..... | 81 |
| Figura 28. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la nueva medida del tiempo de movilidad viajado (TT-R) del grupo C y las otras medidas de movilidad del grupo C (DT-R y GCT-R, tonos de rojo), Así como las tres medidas de los grupos A (azul) y B (tonos de verde)..... | 82 |
| Figura 29. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la nueva medida de las celdas de la cuadrícula de movilidad viajaron (GCT-R) del grupo C y las otras medidas de movilidad del grupo C (DT-R y TT-R,), así como las tres medidas de los grupos A (azul) y B (tonos de verde)..... | 83 |
| Figura 30. UI para la predicción de lugar siguiente individual..... | 86 |
| Figura 31 UI para la predicción agregada del siguiente lugar..... | 87 |
| Figura 32 Una instantánea de una serie de heatmaps que muestran la movilidad de las personas de bukit batok con ct fijado a la medianoche a las 2 AM mientras qt toma valores para cubrir un día entero | 88 |

LISTA DE TABLAS

| | |
|---|----|
| Tabla 1. Resumen de características de varios modelos | 26 |
| Tabla 2. Descripción detallada del proceso de caracterización de la variable..... | 41 |

INTRODUCCIÓN

En la actualidad los requerimientos de movilidad han generado problemas sociales y culturales, en los que se han visto afectadas las personas y las entidades de transporte. Por lo tanto, no se están llevando a cabo soluciones adecuadas para mejorar la movilidad, como una posible solución a esta problemática se tiene la información que se puede extraer de los registros de llamadas de los teléfonos celulares. Hoy en día se están creando grandes volúmenes de datos con estos registros, es un nuevo insumo que no solo puede generar un valor económico sino también bien social.

Las entidades prestadoras de servicio y las instituciones de control no cuentan con herramientas que les ayuden a subsanar los problemas de movilidad, por lo cual se pretende realizar una guía que permita llevar a cabo un análisis de big data con el propósito de obtener los patrones de movilidad a partir de los registros de llamadas de un proveedor de servicios de telecomunicaciones celular.

Esta aproximación ha permitido obtener resultados en otras ciudades, y es lo que se pretende expresar en esta investigación que puede llegar a ser un componente importante en la caracterización de la movilidad de una ciudad o espacio geográfico.

1. GENERALIDADES

1.1. TÍTULO

Análisis Big Data de los patrones de movilidad a partir de registros de llamada.

1.2. FORMULACIÓN DEL PROBLEMA

Los entes encargados de establecer las soluciones de movilidad para una ciudad no cuentan con información suficiente que les permita tomar decisiones que ayuden a mejorar los tiempos de transporte de las personas. De manera adicional, no se cuenta con información centralizada que permita llevar a cabo el análisis de datos con el propósito de generar políticas y lineamientos de movilidad basadas en datos reales.

Las empresas de transporte de pasajeros cuentan con información parcial de movimientos en términos de cargue y descargue de pasajeros, pero no se tienen en cuenta los ciclos de movilidad ocasionados por las personas que se desplazan en vehículos particulares.

1.3. JUSTIFICACIÓN

Como se ha visto en la definición del problema, los ciclos de movilidad no están definidos de una manera clara, limitando la posibilidad de solucionar la movilidad de una determinada ciudad. Por lo tanto, la definición de mecanismos que ayuden a reducir esta problemática evitaría la congestión de vehículos, y ayudaría a definir mejores rutas de desplazamientos hacia un lugar específico. Por esto se hace necesario la implementación de un estudio de los patrones de movilidad con base en la información que se puede extraer de los registros de llamadas realizadas en la red celular como una nueva alternativa a mejorar la movilidad.

En Colombia no se ha realizado un estudio de este tipo que ayude a mejorar la movilidad teniendo en cuenta esta nueva alternativa, por lo que sería un gran aporte para las entidades de transporte y gubernamentales de contar con un estudio que

le permita tener apoyo en la implementación de nuevas rutas o carreteras para mejorar la movilidad.

Durante varios años, ha habido un aumento dramático en el número de usuarios con dispositivos móviles en todo el mundo. Se ha informado de que, para el año 2020 habrá más de 50 millones de dispositivos conectados [7]. Por este motivo este análisis podrá ser una primera aproximación para atacar a la problemática de la movilidad utilizando los registros de llamadas de una red celular.

1.4. OBJETIVOS

1.4.1. OBJETIVO GENERAL

Definir los requerimientos para hacer un análisis de Big Data que permitan obtener los patrones de movilidad de una ciudad.

1.4.2. OBJETIVOS ESPECIFICOS

- Realizar un estudio de estado del arte de los algoritmos existentes que pueden ser aplicados a extraer información de registros de llamada en una red celular.
- Realizar una caracterización de todas las variables de los registros llamadas, para así identificar cuáles son de ayuda para poder realizar el proceso de análisis de movilidad.
- Establecer una guía paso a paso que integre procesos como curación de datos, análisis de variables ya identificadas y la obtención de resultados para visualización de los datos.

1.5. MARCO REFERENCIAL

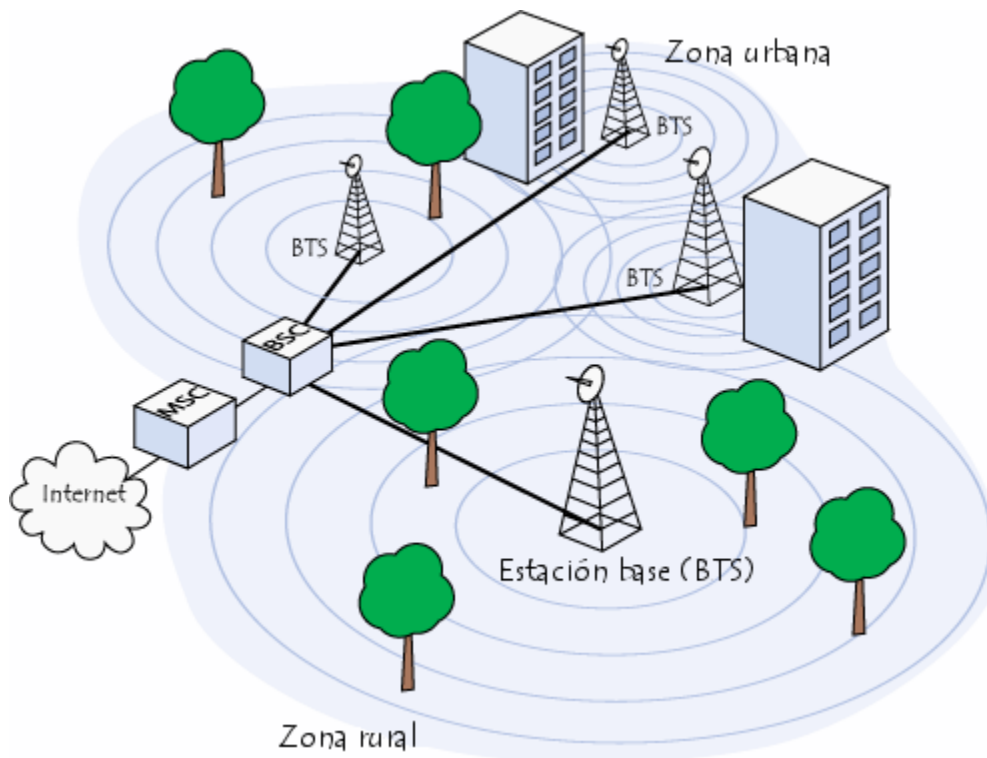
1.5.1. MARCO TEORICO

La gran necesidad que se tiene para solucionar problemas de movilidad en la actualidad es de gran impacto a nivel social, y no se ha llevado un proceso riguroso que solucione esta situación.

Dado que el tema central del análisis estará puesto en mejorar problemas de movilidad a través de la información que se puede extraer de los registros de llamadas de teléfonos celulares. Esto se hace con base en los datos generados a través de un sistema de red celular, será necesario plantear algunos conceptos y teorías que sirvan de ejes conceptuales que se usaran en el desarrollo del proyecto.

- Sistema de red celular

Figura 1. Funcionamiento de un sistema celular



Fuente: <http://es.ccm.net/contents/681-estandar-gsm-sistema-global-de-comunicaciones-moviles>

Las estaciones base de una red celular están conectadas a un controlador de estaciones base (o BSC), que administra la distribución de los recursos. El sistema compuesto del controlador de estaciones base y sus estaciones base conectadas es el subsistema de estaciones base (o BSS).

Los controladores de estaciones base están físicamente conectados al centro de conmutación móvil (MSC) que los conecta con la red de telefonía pública y con Internet; lo administra el operador de la red telefónica. El MSC pertenece a un subsistema de conmutación de red (NSS) que gestiona las identidades de los usuarios, su ubicación y el establecimiento de comunicaciones con otros usuarios.

Generalmente, el MSC se conecta a bases de datos que proporcionan funciones adicionales:

- **El Registro de ubicación de origen (HLR):** Es una base de datos que contiene información (posición geográfica, información administrativa, etc.) de los abonados registrados dentro de la zona del conmutador (MSC).
- **El Registro de ubicación de visitante (VLR):** Es una base de datos que contiene información de usuarios que no son abonados locales. El VLR recupera los datos de un usuario nuevo del HLR de la zona de abonado del usuario. Los datos se conservan mientras el usuario está dentro de la zona y se eliminan en cuanto abandona la zona o después de un período de inactividad prolongado (terminal apagada).
- **El Registro de identificación del equipo (EIR):** Es una base de datos que contiene la lista de terminales móviles.
- **El Centro de autenticación (AUC):** Verifica las identidades de los usuarios.

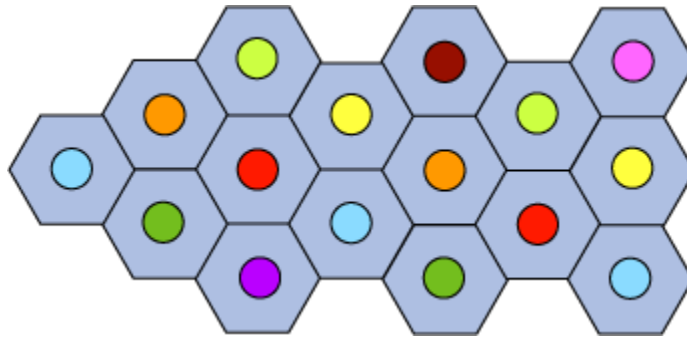
La red celular compuesta de esta manera está diseñada para admitir movilidad a través de la gestión de traspasos (movimientos que se realizan de una celda a otra).

[1]

- Celda

Las redes de telefonía móvil se basan en el concepto de celdas, es decir zonas circulares que se superponen para cubrir un área geográfica.

Figura 2. Representación de una celda



Fuente: <http://es.ccm.net/contents/681-estandar-gsm-sistema-global-de-comunicaciones-moviles>

Las redes celulares se basan en el uso de un transmisor-receptor central en cada celda, denominado "estación base" (o Estación base transceptora, BTS).

Cuanto menor sea el radio de una celda, mayor será el ancho de banda disponible. Por lo tanto, en zonas urbanas muy pobladas, hay celdas con un radio de unos cientos de metros, mientras que en zonas rurales hay celdas enormes de hasta 30 kilómetros que proporcionan cobertura.

En una red celular, cada celda está rodeada por 6 celdas contiguas (por esto las celdas generalmente se dibujan como un hexágono). Para evitar interferencia, las celdas adyacentes no pueden usar la misma frecuencia. En la práctica, dos celdas que usan el mismo rango de frecuencia deben estar separadas por una distancia equivalente a dos o tres veces el diámetro de la celda. [1]

- Celdas urbanas y rurales

El tamaño de las celdas depende de muchos factores como el tipo de antenas utilizado, el terreno (llanuras, montañas, valles, etc.), la ubicación de la instalación (área rural, urbana, etc.), la densidad de población, etc. El tamaño de la celda está también limitado por el alcance del teléfono móvil que debe ser capaz de establecer el enlace de retorno.

Además, una estación base tiene una capacidad de transmisión limitada y sólo puede gestionar simultáneamente un determinado número de llamadas.

Por ello, en las zonas urbanas, con alta densidad de población y un número importante de comunicaciones, las celdas tienden a ser numerosas y pequeñas (a cientos o incluso a sólo unas decenas de metros de distancia).

En las zonas rurales, con menor densidad de población, el tamaño de las celdas es mucho mayor, a veces, hasta varios kilómetros, aunque rara vez más de diez kilómetros.

Es importante subrayar que la disminución de la potencia de la señal emitida por las antenas conlleva una reducción de la cobertura de las celdas.

Al contrario, el incremento del número de celdas mejora la capacidad de transmisión de tráfico de voz o datos de la red, pero requiere que se aumente el número de estaciones base [2].

Figura 3. Localización de antenas rural y urbana



Fuente: <http://www.ondes-radio.orange.com/es/redes-y-estaciones-base/como-funciona-una-red-movil>

- **Radio base**

La radio base o estación base (RBS por sus siglas en inglés o EB), está siempre en conexión con la MTSO (Mobile Telephone Switching Office) o MSC (Mobile Switching Central) por medio de enlaces punto a punto. Esta es considerada como una estación de relevo para señales con información de datos y voz. La RBS supervisa la calidad de la transmisión durante una llamada en progreso por medio de frecuencias de supervisión y la medición de la intensidad de la potencia de las llamadas recibidas [3].

- **CDR (Call Detail Record)**

Es un registro de todos los pasos que concurren en una llamada, pero con un nivel de detalle muy reducido, cada vez que una persona hace una llamada se envía un mensaje de texto y se genera un registro de llamadas (CDR- Call Detail Record), que registra la hora, el día, el tiempo de duración, el tipo de comunicación, y una amplificación de la torre celular que maneja la solicitud. La ubicación aproximada de

un teléfono móvil y su usuario se puede reconstruir con los CDRs asociados al teléfono con los lugares de las torres celulares que manejan las llamadas.

- Importancia de los CDR

Los registros de telefonía celular se han convertido en una forma de estudiar la movilidad de las personas, estos registros se recogen continuamente por los proveedores del servicio celular para mejorar el servicio de las redes celular.

Cada vez que una persona hace una llamada se envía un mensaje de texto y se genera un registro de llamadas (CDR- Call Detail Record), que registra la hora, el día, el tiempo de duración, el tipo de comunicación, y una amplificación de la torre celular que maneja la solicitud. La ubicación aproximada de un teléfono móvil y su usuario se puede reconstruir con los CDRs asociados al teléfono con los lugares (latitud y longitud) de las torres celulares que manejan las llamadas, con estos datos permiten hacer un estudio de la movilidad de las personas, ya que el uso de los teléfonos móvil en las personas aumenta drásticamente en todo el mundo.

- Comunicación de la radio base con los celulares

Las comunicaciones móviles siguen el principio general de la telefonía: Conectar dos usuarios remotos a través del equipo de red de un operador responsable de la gestión del servicio. Sin embargo, a diferencia de los teléfonos fijos, en la red móvil no existen pares de cobre ni fibra óptica, y las transmisiones de radio constituyen el enlace final. El teléfono móvil del usuario comunica a través del aire con una antena, que a su vez comunica con la central del operador. Ésta encamina la comunicación hacia la parte correspondiente en la red fija o a través de otras antenas.

Para que la comunicación sea efectiva, el usuario móvil debe estar en el área de alcance de una antena. Ésta tiene un alcance limitado y cubre una pequeña área alrededor, llamada "celda" (de ahí el otro nombre de "red de celdas" o "red celular" utilizado a menudo para designar las redes móviles). Para cubrir el máximo territorio y garantizar que los usuarios puedan siempre llamar, los operadores despliegan

miles de celdas, cada una equipada con estaciones base, asegurándose de que no haya huecos entre ellas para que nunca se pierda la localización de los usuarios [2].

Conceptos de procesamiento de datos

- **Big Data**

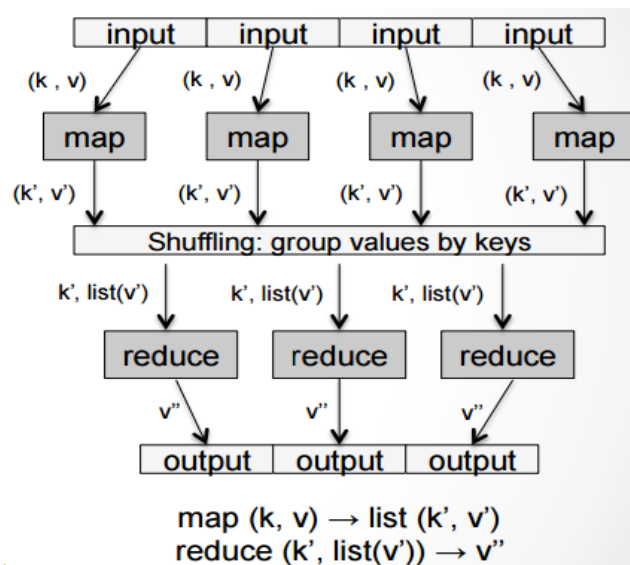
Big data es una colección de datos grande, complejos, muy difícil de procesar a través de herramientas de gestión y procesamiento de datos tradicionales.

Las agencias de tráfico estatales y regionales, y los responsables de municipios y distritos (instituciones públicas) son los principales interesados en las ventajas que puede aportar Big Data para planificar, gestionar y regular la movilidad en las zonas urbanas. [4]

- **Algoritmo que permite hacer análisis de Big Data [5]**

- **MapReduce:** MapReduce es el entorno más popular para Big Data
 - Basado en la estructura {clave, valor}
 - Dos operaciones:
 1. Función Map: Procesa bloques de información
 2. Función Reduce: Fusiona los resultados previos de acuerdo con su clave.
 - Una etapa intermedia de agrupamiento por clave (Shuffling)

Figura 4. Funcionamiento del algoritmo MapReduce



Fuente: [http://madm.uib.es/wp-content/uploads/2016/06/Jose-Manuel-Benitez-Sanchez-Big-Data-Algoritmos tecnologia y aplicaciones.pdf](http://madm.uib.es/wp-content/uploads/2016/06/Jose-Manuel-Benitez-Sanchez-Big-Data-Algoritmos_tecnologia_y_aplicaciones.pdf).

- **Hadoop:** Es una implementación de código abierto del paradigma de programación MapReduce.
- **Spark:** Entorno de trabajo (gestión de ejecución, API) genérico y rápido (hasta 100 veces más que Hadoop) para procesamiento de datos masivos.
 - Big Data “in-memory”, Spark permite realizar trabajos paralelizados totalmente en memoria:
 - Reducción de tiempos
 - Procesos iterativos
 - Esquema de computación más flexible que MapReduce.
 - Permite los flujos a cíclicos de procesamiento de datos.
- **Flink:** Se trata de un motor de procesamiento de streams o flujos de datos que proporciona capacidades de distribución de datos, comunicaciones y, muy importante, tolerancia a fallos a las computaciones.

1.6. HIPOTESIS

¿Será posible que, mediante un análisis de las variables contenidas en los registros de llamadas vía celular, establecer los requerimientos para la realización de un análisis de Big Data que permita establecer los flujos de movilidad de una ciudad?

Con esta pregunta, es con la cual se quiere hacer la investigación en este proyecto, y llegar a una respuesta clara que permita tomar decisiones frente a la problemática asociada a la movilidad.

1.7. TIPO DE INVESTIGACIÓN

La investigación es de tipo cuantitativa

1.8. METODOLOGIA

1.8.1. UBICACIÓN DEL EXPERIMENTO

Es una investigación que se va generar en términos teóricos, en el que se tomara como referencia a una determinada ciudad que cuente con los datos requeridos de los registros de llamadas para realizar el análisis.

Ya que la investigación será de tipo cuantitativo se terminará midiendo en término de valores que sean medibles, ya sea a través de fórmulas o de resultados obtenidos.

1.8.2. DISEÑO EXPERIMENTAL

Se realizará a través de un escenario de validación de existencia y calidad

Validación de Calidad: Se deberá definir cuáles son los parámetros o variables de calidad que va definir si la guía cumple o no con ciertos tipos de expectativas, para concluir si se cumplió con el objetivo.

Por ejemplo, condiciones que debe cumplir en términos de:

- Presentación

- Índice
- Justificado
- Orden coherente
- Referencias bibliográficas

➤ Contenido y Alcance

- Estudio de algoritmos existentes y cuales se van a emplear
- Proceso de caracterización de las variables que se van a validar.

1.8.3. SELECCIÓN DE LA MUESTRA/ DESCRIPCIÓN DE UNIDADES EXPERIMENTALES

En la selección de las muestras se utilizará un método de muestreo no probabilístico o muestreo estratificado, en el cual permita realizar el procedimiento a través de diferentes tipos de algoritmos que permitan identificar las variables que ayuden a realizar el análisis.

1.8.4. RECOLECCIÓN DE DATOS

Esta parte no se realizará, ya que se asume que los datos ya se tienen.

1.8.5. PROCEDIMIENTOS PARA SU EVALUACIÓN

Etapas I:

- Recolección de la información sobre los procesos realizados actualmente, relacionados con el tema.
- Realizar la planeación del proyecto.
- Especificar los requerimientos de la investigación.

Etapas II:

- Investigación de los diferentes tipos de algoritmos, en los cuales permitan realizar la identificación de las variables que se requieren de los registros de llamadas.

Etapas III:

- Definir tipos de pruebas que se puedan aplicar en la investigación.
- **Pruebas piloto:**
 - Poner a prueba diferentes aspectos de la ejecución de la investigación.
 - Evaluar aspectos técnicos, mejorar y corregir todas las deficiencias que se encuentren.
 - Comprobar la viabilidad del proyecto.

Etapas IV:

- Analizar los resultados que se obtuvieron de acuerdo con las técnicas que se eligieron, para realizar el proceso curación de datos de los resultados, para concluir y dar recomendaciones respecto a la investigación realizada.

2. ESTUDIO DEL ESTADO DE ARTE DE LOS ALGORITMOS EXISTENTES PARA EXTRAER INFORMACIÓN DE LOS REGISTROS DE LLAMADA EN UNA RED CELULAR.

En este proyecto se realiza un proceso de cómo obtener patrones de movilidad de un sector de una ciudad a través de la información que se puede extraer de los registros de llamadas, se realizará un estudio de estado del arte de diferentes algoritmos que permita extraer esta información de una red celular, y conocer el comportamiento de una persona que realiza una llamada cuando se enlaza a una red celular.

Los algoritmos describen una serie de pasos y de cálculos que, de acuerdo a una información suministrada, podrá obtener una solución respecto a una situación que se quiera analizar.

2.1.ALGORITMOS

Los algoritmos existentes pueden ser aplicados de acuerdo a su funcionamiento para realizar el cálculo con base a radio bases que se encuentran alrededor, para determinar el trayecto de una persona que realiza una llamada en un punto específico de una ciudad, la llamada se enlaza a una radio base más cercana que se encuentre en la zona, en la que pueden ocurrir dos circunstancias.

1. Que la persona realice la llamada desde una parte estática.
2. Que la persona realice la llamada en movimiento.

Para esta investigación servirán los registros de llamadas que se realicen en movimiento, ya que se comunicará con diferentes radio bases permitiendo conocer los registros generados en cada radio base que se conecta y realizar el cálculo del viaje cuando la persona llega a su destino, con el fin de conocer su comportamiento.

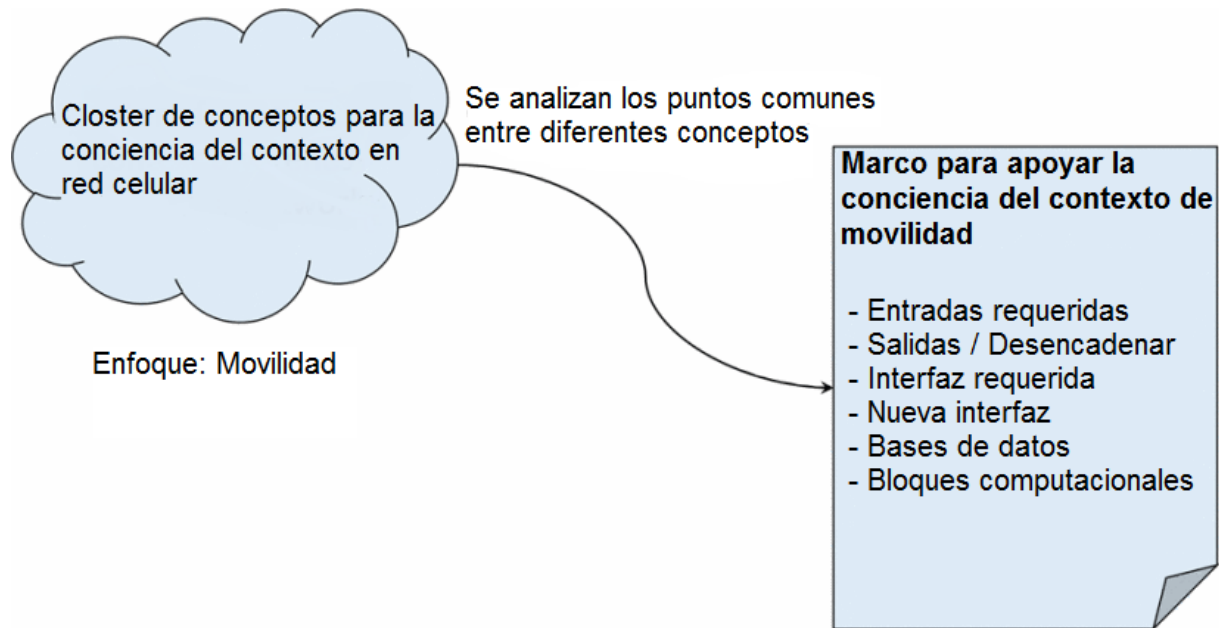
2.1.1. MARCO PARA APOYAR EL CONOCIMIENTO DEL CONTEXTO DE MOVILIDAD EN REDES CELULARES [6]

La comunicación móvil se ha convertido en una tecnología ampliamente utilizada en el día a día, con un horizonte de expansión de sus aplicaciones desde su creación. La comunicación móvil ha evolucionado a través de varias generaciones de tecnología y actualmente está en camino hacia la quinta generación (5G). Durante varios años, ha habido un aumento dramático en el número de usuarios móviles en todo el mundo. Se ha informado que para el año 2020 habrá más de 50 millones de dispositivos conectados. Además, los servicios de banda ancha móvil están creciendo en popularidad junto con el apoyo a mano y dispositivos personales. Estos factores han dado lugar a una avalancha de crecimiento del volumen de tráfico y se prevé que el volumen de tráfico móvil será por lo menos 1000 veces mayor en el año 2020 que el actual. Acomodar un número tan grande de usuarios y el volumen de tráfico resultante es un desafío clave en 5G. Proporcionar una calidad de servicio uniforme ininterrumpida y mejor QoE independientemente de la alta movilidad y el alto volumen de tráfico son las principales preocupaciones que se abordarán en 5G.

La información contextual es la información que permite la percepción de estados y situaciones de entidades de red (por ejemplo, nodos de red, terminales, usuarios, etc.) y sus relaciones interactivas. La red de radio se convierte en contexto consciente cuando se utiliza información de contexto para ayudar y optimizar el funcionamiento de la red. El uso de la información del contexto y la construcción de la conciencia del contexto se estudió por primera vez en el campo de la informática omnipresente. La información pertinente, como la identidad del usuario, la ubicación, etc., se recogieron del entorno y se alimentaron al sistema informático. Basándose en esta información, el sistema se adaptaría en consecuencia para optimizar su rendimiento. En los últimos tiempos, los

sistemas móviles e inalámbricos han comenzado a explotar los conceptos de conciencia de contexto, haciendo que la red de radio pueda adaptar sus configuraciones o acciones al cambiante entorno físico y de aplicaciones. El conocimiento del contexto en las redes celulares se reconoce como un pilar importante en la solución de los desafíos clave en la comunicación móvil 5G.

Figura 5. Motivación



Fuente: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/7504249/>.

El estudio de este trabajo se ocupa específicamente de la conciencia del contexto en las redes celulares, centrado en la movilidad de los usuarios. La predicción de la movilidad ha sido un componente clave en la construcción de la conciencia del contexto. Al anticipar / aprender el comportamiento de movilidad del usuario, es posible diseñar varios esquemas de gestión de recursos conscientes del contexto, procedimientos de traspaso, sistemas de activación / desactivación de células, etc. En la literatura existen varios esquemas relacionados que explotan el comportamiento de movilidad de usuarios para crear conciencia de contexto. Este trabajo propone un marco para apoyar la conciencia de contexto de movilidad en la red celular analizando los puntos comunes entre varios esquemas conscientes del contexto. Este trabajo también investiga sobre el conjunto de

entradas requeridas, señalización e interfaces. Además, se diseña un esquema de asignación de recursos consciente del contexto que hace uso de nuevas interfaces e información que surgen de la infraestructura vehicular. Los resultados de la simulación muestran mejoras sustanciales en el rendimiento del usuario, apoyando así la necesidad de nuevas interfaces.

2.1.2. DISEÑO Y OPTIMIZACIÓN DE RUTAS Y FRECUENCIAS EN EL TRANSPORTE COLECTIVO URBANO, MODELOS Y ALGORITMOS. [7]

La planificación del transporte público urbano colectivo (TPUC) basada en herramientas de apoyo a la decisión cobra cada vez más importancia, tanto en los países desarrollados como en los en vías de desarrollo. Una proporción importante de los viajes en las ciudades medianas y grandes son efectuados utilizando transporte público colectivo. Problemas como la asignación de flota y personal, han recibido amplio tratamiento, contándose con modelos de optimización para los cuales se dispone de algoritmos eficientes de resolución. En cambio, el problema de optimización de rutas y frecuencias posee varias fuentes de complejidad (no linealidad, no convexidad, múltiples objetivos) que dificultan tanto su formulación como la derivación de algoritmos eficientes de resolución. En este trabajo se tratan modelos y algoritmos para el problema de optimización de rutas y frecuencias, seleccionados en el contexto del desarrollo de una herramienta de apoyo a la planificación del sistema de TPUC para la ciudad de Montevideo, Uruguay (1.500.000 habitantes). Todos los trabajos resuelven el problema utilizando algoritmos heurísticos, encontrándose varias aplicaciones de la técnica algoritmos genéticos.

Modelos para la optimización de recorridos y frecuencias de buses

El principal componente que caracteriza a cada uno de los modelos es su formulación. En particular la función objetivo reflejará tanto los intereses de los usuarios (pasajeros) como de los operadores (empresas de transporte). Los modelos presentados en general, buscan maximizar el nivel de servicio, minimizando el uso de los recursos, según determinadas restricciones. Estos objetivos son generalmente contrapuestos, una mejora

en uno implica un detrimento en el otro; la importancia relativa de los componentes de la función objetivo es una decisión política, por tanto, será definida por las entidades reguladoras del sistema. La tabla presenta un resumen de las características de los modelos seleccionados para este trabajo.

Tabla 1. Resumen de características de varios modelos

| <i>Autor(es)</i> | <i>Función objetivo</i> | <i>Restricciones</i> | <i>Aportes</i> | <i>Limitaciones</i> |
|---|---|---|--|---|
| Baaj y Mahmassani (1991) | Min. tiempos de transferencia y tamaño de flota | Frecuencia factible Factor de carga Tamaño de flota | Varios parámetros configurables | Coefficientes de conversión en fn. Objetivo |
| Israeli y Ceder (1993) | Min. tiempos de transferencia y tamaño de flota (multiobjetivo) | No especificadas | Formulación multiobjetivo | |
| Ngamchai y Lovell (2000) | Min. tiempos de transferencia y tamaño de flota (detallado) | Factor de carga | Modelo detallado, frecuencias óptimas | Coefficientes de conversión en fn. objetivo |
| Gruttner, Pinninghoff, Tudela y Diaz (2002) | Max. beneficios de operador y min. costos de usuario | Dist. de acceso y egreso (a origen y destino) | Modelo alternativo de asignación (logit) | Falta tratamiento de frecuencias y flota Coefficientes de conversión en fn. objetivo |

Fuente: <https://www.fing.edu.uy/inco/pedeciba/bibliote/reptec/TR0307.pdf>

Baaj y Mahmassani (1991)

Se plantea minimizar los tiempos totales de transferencia de pasajeros y el tamaño de la flota requerida, sujeto a restricciones de frecuencia, factor de carga y tamaño de flota. La formulación del modelo es:

$$\min \left\{ C_1 \sum_{i=1}^n \sum_{j=1}^n d_{ij} t_{ij} + C_2 \sum_{k \in R} f_k t_k \right\}$$

sa

$$f_k \geq f_{\min} \quad \forall \quad k \in R \quad (\text{frecuencia factible})$$

$$LF_k = \frac{(Q_k)_{\max}}{f_k CAP} \leq LF_{\max} \quad \forall \quad k \in R \quad (\text{factor de carga})$$

$$\sum_{k \in R} N_k = \sum_{k \in R} f_k t_k \leq W \quad (\text{tamaño de flota})$$

donde

n : cantidad de nodos de la red;

d_{ij} : demanda (cantidad de viajes por unidad de tiempo) entre los nodos i y j ;

t_{ij} : tiempo total de viaje entre i y j (en vehículo, espera y transferencia, si existe);

N_k : cantidad de buses operando en la ruta k , $N_k = f_k T_k$;

t_{ij} : frecuencia de buses operando en la ruta k ;

f_{\min} : mínima frecuencia de buses permitida para toda ruta;

T_k : tiempo total de viaje de la ruta k ;

W : tamaño de la flota disponible (cantidad de buses por hora);

LF_k : factor de carga en la ruta k ;

$(Q_k)_{\max}$: máximo flujo por arco en la ruta k ;

CAP : capacidad de pasajeros sentados en los buses;

LF_{\max} : máximo factor de carga permitido;

R : conjunto de rutas para una solución dada;

C_1 y C_2 : factores de conversión y pesos relativos de los términos de la función objetivo.

Los principales aspectos del problema son tenidos en cuenta, así como una variedad de parámetros y restricciones (factor de carga, por ejemplo). Es flexible, ya que permite la incorporación del conocimiento de los usuarios, por ejemplo, restricciones de mínima proporción de demanda cubierta en base a viajes sin transferencias o con al menos una transferencia pueden ser agregados al momento de aplicar un método de resolución. Los componentes de la función objetivo se expresan en distintas unidades, obligando a utilizar coeficientes de conversión.

Israeli y Ceder (1993)

Este modelo es similar al propuesto por Baaj y Mahmassani (1991), pero se formula como un problema de optimización multiobjetivo.

$$\begin{aligned} \min \quad & Z_1 = a_1 \sum_{i,j \in N} PH_{ij} + a_2 \sum_{i,j \in N} WH_{ij} + a_3 \sum_{r \in R} EH_r \\ \min \quad & Z_2 = FS \end{aligned}$$

PH_{ij} : cantidad de pasajeros/hora, entre los nodos i y j (mide el tiempo de viaje en vehículo de los pasajeros);

WH_{ij} : tiempo de espera de pasajeros entre los nodos i y j ;

EH_r : tiempo de viaje vacío, que refleja la utilización de los buses;

FS : tamaño de la flota;

R : conjunto de rutas para una solución dada;

a_1 , a_2 y a_3 : pesos que reflejan la importancia relativa de los términos de la función Z_1 .

Ngamchai y Lovell (2000)

Con una formulación similar a la propuesta por Baaj y Mahmassani (1991), este modelo permite calcular frecuencias de rutas; aunque requiere del uso de coeficientes de conversión a la misma unidad (\$/hora) de todas las componentes de la función objetivo.

$$\min \{FC + UVC + UWC\}$$

donde

$$FC = \frac{2C_V}{V} \sum_{k=1}^R \frac{d_k}{h_k} \quad (\text{costo de la flota})$$

$$UVC = \frac{\gamma_V}{V} \sum_{i=1}^m \sum_{j=1}^m q_{ij} D_{ij} \quad (\text{costo de viaje en vehículo de los usuarios})$$

$$UWC = \frac{\gamma_w}{2} \sum_{k=1}^R \sum_{i=1}^m \sum_{j=1}^m q_{ij} \alpha_{ijk} h_k \quad (\text{costo de espera de los usuarios})$$

con

m : cantidad de nodos de la red;

R : cantidad de rutas de una solución determinada;

C_V : costo por hora de operación de los buses;

V : velocidad de los buses en la red;

d_k : largo de la ruta k ;

q_{ij} : demanda entre los nodos i y j (cantidad de viajes por hora);

D_{ij} : largo de la ruta más corta seleccionada por los pasajeros viajando de i a j ;

α_{ijk} : $\alpha_{ijk} = 1$ si la ruta k utiliza el arco (i,j) , $\alpha_{ijk} = 0$ en caso contrario;

γ_v y γ_w : coeficientes que reflejan el valor subjetivo de los tiempos de viaje y espera;

h_k : espaciamiento temporal del servicio operante en la ruta k (inverso de la frecuencia),

$$h_k = \min \left(\sqrt{\frac{4d_k C_v}{\gamma_w V \sum_{i=1}^m \sum_{j=1}^m q_{ij} \alpha_{ijk}}}, h_k^{\max} \right)$$

donde h_k^{\max} depende del factor de carga y del arco con mayor flujo en la ruta k .

Gruttner, Pinninghoff, Tudela y Díaz (2002)

Este modelo difiere de todos los anteriores en la especificación de los componentes del sistema. Propone un modelo de asignación alternativo, que usa el método logit mediante el cálculo de utilidades de cada línea para cada par origen-destino (i,j) . No se contemplan aspectos tales como la determinación de frecuencias y dimensionamiento de flota; requiere la utilización de coeficientes de conversión y de valores subjetivos del tiempo. La función objetivo del modelo es:

$$\max \{ \alpha FO(R_i) - \beta FU(R_i) \}$$

donde

R_i : i -ésima ruta válida ($R_i \in R$, conjunto de rutas válidas);

α y β : coeficientes que representan la importancia relativa de cada objetivo;

$FO = IO_L - CO_L$; (función de beneficio del operador)

$IO_L = AF_L T_L$ (ingreso operador)

$CO_L = Distancia_L K_L$ (costo operador)

AF_L : afluencia total de viajes que atrae la ruta L ;

T_L : tarifa cobrada por la línea L ;

K_L : costo unitario de operación por kilómetro;

$FU = CU_L = \sum_i \sum_j (\delta t_{ijL}^a + t_{ijL}^v + \eta t_{ijL}^e) \times VST \times V_{ijL}$ (función de costo del usuario)

$t_{ijL}^a, t_{ijL}^v, t_{ijL}^e$: los tiempos de acceso a la línea, de viaje y de espera respectivamente;

VST : valor subjetivo del tiempo;

V_{ijL} : número de viaje entre cada par origen-destino (i, j) que utilizan la línea L ;

δ y η : pesos relativos de los tiempos de acceso y espera con respecto al tiempo de viaje.

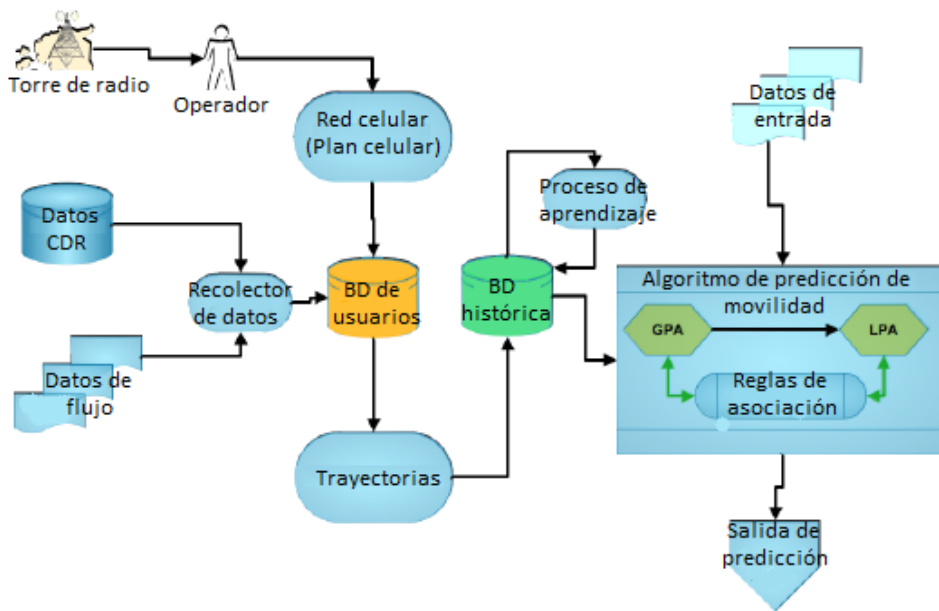
2.1.3. LA PREDICCIÓN DE LA MOVILIDAD DE LOS SUSCRIPTORES DE TELÉFONOS CELULARES USANDO EL ALGORITMO MEJORADO DE LA CADENA DE MARKOV [8]

En este artículo se presenta un método de predicción de la movilidad de los usuarios de teléfonos móviles basados en un algoritmo mejorado de la cadena de Márkov. Los datos de teléfono móvil tienen un carácter altamente dinámico y un aspecto escasamente muestreado; Por lo tanto, la predicción de la ubicación movilidad del usuario plantea un desafío. La mejora puede resumirse como una asociación incrustada de las reglas aplicadas a un algoritmo de cadena de Márkov. La solución propuesta es alentadora para la próxima generación de redes de telefonía móvil y se puede utilizar para optimizar la infraestructura de red móvil, el tráfico de carretera existente, sistemas y localización de seguimiento. La validación del sistema se lleva a cabo utilizando datos reales recogidos desde el campo.

Arquitectura del sistema

La arquitectura del sistema se resume en la Figura 6. Está diseñado para procesar los datos recopilados y también para entrenar el modelo basado en las trayectorias de los usuarios antes de hacer predicciones de la siguiente ubicación de celda del usuario.

Figura 6. Visión general de la arquitectura del sistema



Fuente: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/6856442/>

En primer lugar, los datos se recopilan en la base de datos de usuarios. Luego, se procesa para extraer todas las trayectorias de todos los usuarios y crear lo que se define como trayectorias de usuarios. Estas trayectorias son un conjunto de sucesos posteriores en la entrada y algunas veces pueden pertenecer a la misma celda, lo cual no es relevante para la trayectoria del usuario.

Tales eventos se combinan en una visita celular (Tiempo que tarda un usuario conectado a una torre celular) considerando que el tiempo entre dos eventos no es más largo que una constante específica, tomada como 65 minutos debido a eventos de servicio por hora. Cada visita tiene además de la célula c_i también comienza el tiempo b_i y el tiempo de fin e_i que corresponden al primer y último evento de la secuencia combinada. Por lo tanto, si sólo hay un evento, entonces los tiempos son iguales $b_i = e_i$.

Además, el sistema contiene tres filtros de trayectoria que pueden resultar en la división de la trayectoria en muchas sustracciones (cadenas). Todas las cadenas comprenden la

trayectoria del usuario, pero no hay transición entre las visitas de borde de las cadenas subsiguientes que se considera en los algoritmos.

Los filtros utilizados en el algoritmo son los siguientes:

- El filtro de tiempo sólo define el período y los días de la semana para una trayectoria del usuario.
- El filtro Ping pong handover es un filtrado de acuerdo con los fenómenos ping-pong handover en las trayectorias de los usuarios. Este filtrado se realiza para reducir los errores que se pueden encontrar en el proceso de predicción. El handover ping-pong es cuando los usuarios cambian aleatoriamente y varias veces entre dos células vecinas debido a las diversas condiciones de la propagación de la señal. El filtrado comienza dividiendo la trayectoria donde tres células secuenciales no son únicas, por ejemplo, c_a, c_b, c_a , or c_a, c_b, c_b . La célula media se incluye en ambas cadenas a menos que el triple siguiente también deba dividirse. Esto puede crear visitas duplicadas, pero no afecta los resultados.
- El filtro de separación divide la trayectoria donde el retraso entre visitas posteriores $b_i - e_{i-1}$ es mayor que 2 horas.

Finalmente, después de este proceso se lanza el algoritmo de Márkov mejorado para producir la predicción de movilidad. El algoritmo de predicción contiene dos reglas de asociación (la regla temporal y la regla de comportamiento universal), que se ejecuta por separado en las pruebas. Se hace con el fin de tener una idea de cuál de las reglas de asociación da mejores resultados en el proceso de predicción.

Algoritmo para predecir la movilidad

El hecho de que el comportamiento humano sea repetitivo lleva a suponer que es fácil predecir la movilidad de los usuarios; Sin embargo, los usuarios a veces pueden ser impredecibles en su movimiento cuando cambian sus hábitos. Por esta razón se introducen dos reglas de asociación para tener en cuenta esta irregularidad. Primero, se comienza introduciendo el enfoque de Márkov clásico, entonces se introduce el realce (la regla de comportamiento universal y la regla temporal). En general, los enfoques de la familia Márkov de orden k suponen que la predicción de la siguiente ubicación puede

hacerse utilizando la ubicación o contexto actual. Lo que significa que se está buscando las k ubicaciones más recientes en el historial de trayectoria T . El modelo de Márkov tiene dos componentes que lo representan: Cada estado o evento es un contexto y las transiciones son las ubicaciones posibles después de ese contexto. Si se considera la trayectoria histórica de un usuario $T = c_1 \rightarrow c_2 \rightarrow \dots \rightarrow c_n$ y la trayectoria aleatoria subsecuencia $T(i, j) = c_i \rightarrow c_{i+1} \rightarrow \dots \rightarrow c_j$ para cualquier $1 \leq i \leq j \leq n$. Entonces se toma X , una variable aleatoria que ilustra la ubicación del usuario.

Definición 1: Φ es el conjunto de todos los usuarios de la base de datos.

Definición 2: Contexto $\gamma = T(n - k + 1, n)$ es el contexto en un momento específico.

Por lo tanto, X se comporta de la siguiente manera:

$$\mathbf{P}(X_{n+1}|X_{1:n}) = \mathbf{P}(X_{n+1}|X_{n-k+1:n}) \quad \text{Ec. 1, Cap. 2.1.3}$$

$$\mathbf{P}(X_{n+1}|X_{n-k+1:n}) = \mathbf{P}(X_{i+k+1}|X_{i+1,i+k}), \quad \text{Ec. 2, Cap. 2.1.3}$$

Donde, $\mathbf{P}(\dots | \dots)$ es la probabilidad condicional. La ecuación (1) ilustra la dependencia del contexto γ y la ecuación (2) muestra el estado estacionario donde la probabilidad es la misma dondequiera que el contexto sea el mismo. En nuestro caso el orden es $k = 2$ y estas probabilidades representan las probabilidades de transición de una célula a otra. El proceso Márkov tiene dos componentes: el primero es el GPA que está diseñado para los usuarios que existen en la base de datos de formación y LPA es para los usuarios que no existen en la base de datos de formación. Además, tanto el GPA como el LPA utilizan las probabilidades de transición definidas anteriormente.

A. GPA

El enfoque GPA utiliza un proceso de Márkov de segundo orden en el que sus estados discretos son las células de la red celular. Por lo tanto, la probabilidad de transición del GPA para cada usuario móvil de pasar a la siguiente celda c_{n+1} depende no sólo de la celda actual c_n sino también de la anterior c_{n-1} . La probabilidad de transición se define de la siguiente manera:

$$P_g(c_{n+1}|c_{n-1}, c_n) = \frac{M(c_{n-1} \rightarrow c_n, c_{n+1})}{\sum_{c \in \Psi} M(c_{n-1} \rightarrow c_n, c)}, \quad \text{Ec. 3, Cap. 2.1.3}$$

Donde $M(c_{n-1} \rightarrow c_n, c_{n+1})$ es el número de transiciones del usuario móvil de la celda c_n a c_{n+1} en el pasado, sabiendo que el usuario fue antes en c_{n-1} . Esta probabilidad puede fallar si la célula no existe en la base de datos de entrenamiento y será igual a cero. En esta situación, el ACP no puede predecir; Por lo tanto, el enfoque LPA toma el relevo y trata de solucionar este fracaso.

B. LPA

El LPA se utiliza cuando el GPA falla. El enfoque es un primer orden de Márkov basado en la celda actual c_n y la celda siguiente c_{n+1} . En este caso, la probabilidad de transición se expresa de la siguiente manera:

$$P_l(c_{n+1}|c_n) = \frac{N(c_n, c_{n+1})}{\sum_{c \in \Psi} N(c_n, c)}, \quad \text{Ec. 4, Cap. 2.1.3}$$

Donde $N(c_n, c_{n+1})$ es el número de transiciones del usuario móvil desde la celda c_{n-1} a c_n . En las siguientes secciones, se representa dos mejoras que se agregan a los algoritmos GPA y LPA. La primera regla de asociación que se ha añadido es la regla de comportamiento universal de los patrones de movilidad de los usuarios. El nuevo enfoque se llama como el algoritmo de predicción universal global (GUPA) y el algoritmo de predicción universal local (LUPA). La segunda regla, que es la regla temporal, genera el algoritmo de predicción temporal global (GTPA) y el algoritmo de predicción temporal local (LTPA).

C. GUPA

El método GUPA es similar al GPA excepto que tiene en cuenta el comportamiento de movilidad universal de todos los usuarios $v \in \Phi$ en la base de datos histórica. La probabilidad de transición se define por:

$$P_{ug}(c_{n+1}|c_{n-1}, c_n) = \frac{\sum_{\Phi} M_v(c_{n-1} \rightarrow c_n, c_{n+1})}{\sum_{\Phi} \sum_{c \in \Psi} M_v(c_{n-1} \rightarrow c_n, c)}, \quad \text{Ec. 5, Cap. 2.1.3}$$

Donde M_v es el recuento de transiciones triples M para el usuario v

D. LUPA

El enfoque LUPA tiene en cuenta todos los usuarios, lo que significa que la ecuación será:

$$P_{ul}(c_{n+1}|c_n) = \frac{\sum_{\Phi} N_v(c_n, c_{n+1})}{\sum_{\Phi} \sum_{c \in \Psi} N(c_n, c)}, \quad \text{Ec. 6, Cap. 2.1.3}$$

Donde N_v es el número de transiciones N para el usuario v.

E. GTPA

El GTPA utiliza la misma noción que GPA con una restricción en el tiempo. Se define la ventana de tiempo para filtrar las transiciones que ocurren "durante" el intervalo. No hay solución obvia óptima para el problema de filtrado, por lo tanto, se implementa uno que no es demasiado restrictivo, pero todavía ayuda a retener las transiciones de usuario pertinentes. Para la última célula c_{n-1} y c_n en la cadena, para la que se tiene que predecir

la transición, tiene los tiempos de inicio y de fin de las visitas celulares (b_{n-1}, e_{n-1}) y (b_n, e_n)². El intervalo de la ventana es definido como (ver Figura 7):

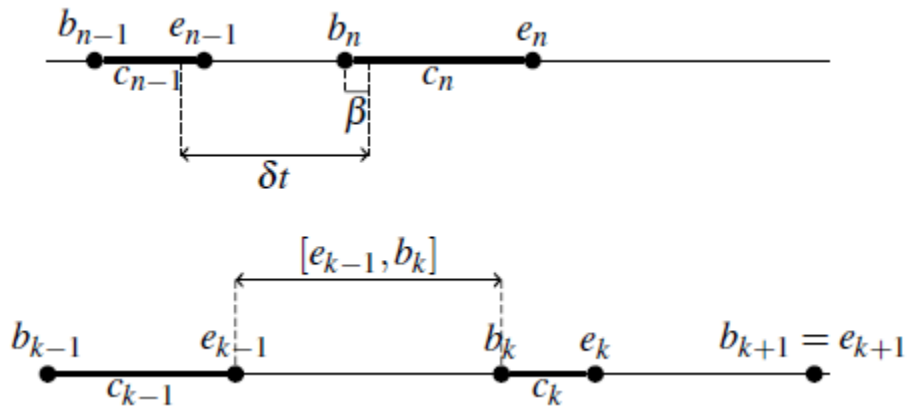
$$\delta t = [e_{n-1} - \beta, b_n + \beta] \quad \text{Ec. 7, Cap. 2.1.3}$$

Donde β es el parámetro de la extensión igual a 15 minutos. El tiempo se define sin tener en cuenta la fecha del evento y su espacio es cíclico, por ejemplo: $[23h; 1h] \cap [0h; 2h] = [0h; 1h]$. A continuación, seleccionamos todos los eventos que sucedieron en el intervalo de tiempo δt y se usa la siguiente fórmula:

$$P_{ig}(c_{n+1}|c_{n-1}, c_n) = \frac{M(c_{n-1} \rightarrow c_n, c_{n+1} | F([e_{k-1}, b_k], \delta t))}{\sum_{c \in \Psi} M(c_{n-1} \rightarrow c_n, c | F([e_{k-1}, b_k], \delta t))} \quad \text{Ec. 8, Cap. 2.1.3}$$

Aquí $M(\dots | C(k))$ es el número de transiciones triples $c_{n-1} \rightarrow c_n \rightarrow c_{n+1}$ filtrado por un predicado C que se refiere a la transición que se está probando a través del índice k . La variable objetivo F garantiza que la superposición de los intervalos no está vacía. $F(d_1, d_2) = d_1 \cap d_2 \neq 0$. Esto deja sólo cadenas triples cuyo tiempo de transición del nodo anterior a la corriente se superpone a δt (Figura 7).

Figura 7. Superposición de intervalos en GTPA



Fuente: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/6856442/>

F. LTPA

El LTPA utiliza el mismo principio que LPA con el intervalo de tiempo δt como se define en el GTPA. Así, las ecuaciones son:

$$P_{il}(c_{n+1}|c_n) = \frac{N(c_n, c_{n+1} | c_{k:k+1} = c_{n:n+1} \wedge F([e_k, b_{k+1}], \delta t))}{\sum_{c \in \Psi} N(c_n, c | F([e_{k-1}, b_k], \delta t))}$$

Donde

Ec. 9, Cap. 2.1.3

$$N(c_n, c_{n+1} | c_{k:k+1} = c_{n:n+1} \wedge F([e_k, b_{k+1}], \delta t))$$

Denota el número de transiciones de los usuarios móviles desde la celda c_{n-1} a c_n con respecto a $F([e_k, b_{k+1}], \delta t)$.

2.1.3.1. SIADDE [9]

Un algoritmo que permite obtener patrones de movilidad utilizando los datos de las tarjetas de transporte de los usuarios.

Este sistema permite a los responsables del medio de transporte conocer para cada tarjeta (un número para ellos) las zonas donde el usuario accede al medio de transporte y aquellas donde se baja (también cuenta con un sensor en esa zona). Esta información, que de forma individual no tiene mayor importancia, puede tener una gran validación si se consigue aprovechar la de todos los usuarios en su conjunto.

Para obtener información útil de un conjunto tan elevado de datos, la empresa Terrain Technologies ha desarrollado un complejo algoritmo, de nombre SIADDE, que va a permitir a los responsables de estos sistemas optimizar la frecuencia de las líneas, combinar eficazmente los medios de transporte y, en definitiva, desarrollar políticas de movilidad teniendo en cuenta los usos reales de los ciudadanos, y adaptadas a diferentes épocas del año o acontecimientos destacados que provocan variaciones.

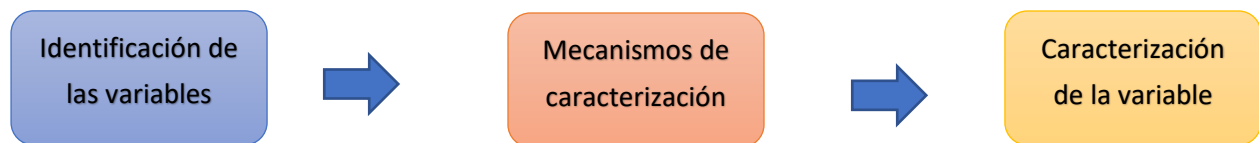
3. CARACTERIZACIÓN DE LAS VARIABLES

En este capítulo se plantea la caracterización de las variables para tener en cuenta en el proceso del análisis de movilidad, en el que se describe el aporte a partir de un estudio previo realizado.

La caracterización de las variables consiste en identificar condiciones y/o elementos que hacen parte del análisis de movilidad, tales como: Nombre de la variable, descripción de la variable, importancia de la variable, aporte de la variable, indicador de la variable, periodo de tiempo, porcentaje de importancia de la variable.

3.1. RUTA PARA LA CARACTERIZACIÓN DE LAS VARIABLES

Figura 8. Proceso de la caracterización



Fuente: Autor

La caracterización consta de varios pasos, primero se identifica las variables que se van a ver involucradas en el análisis, segundo se define los mecanismos de caracterización, tercero se caracteriza la variable, y por último se analiza el resultado de cada variable, que será mirado en detalle en el siguiente capítulo, en el que se integrará cada una de estas variables, para obtener una medición de movilidad de una zona específica de una ciudad.

3.1.1. IDENTIFICACIÓN DE LAS VARIABLES A INVOLUCRAR EN EL ANÁLISIS

Para el proceso de identificación de las variables se describen dos circunstancias claves de la movilidad. La primera es la frecuencia de movimiento, y la segunda es el número

de veces que una persona va alguna parte. Cuanto mayor sea la frecuencia, o más veces una persona se mueve, mayor debe ser el valor de su medida de movilidad. Lo que constituye ir a algún lugar depende de la definición que varían según el estudio y el contexto.

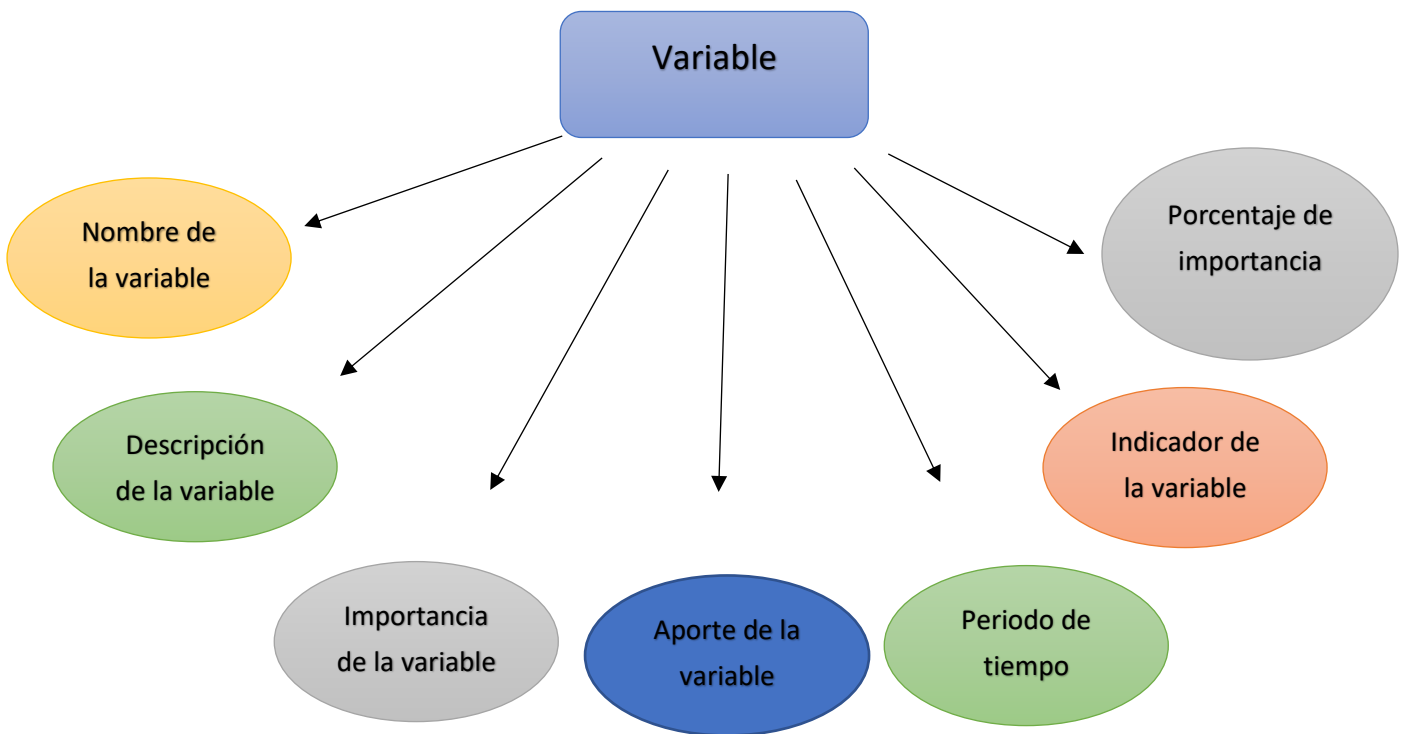
Con base a lo anterior se identifican ciertas variables para ser analizadas:

- CDR (Registros de detalles de llamadas)
- NTU (Numero de torres usadas)
- DT-SL (Distancia recorrida-línea recta)
- MDT (Distancia máxima recorrida)
- ROG (Radio de giro)

3.1.2. DEFINICIÓN DE MECANISMOS DE CARACTERIZACIÓN DE LAS VARIABLES

Las variables se caracterizan teniendo en cuenta las siguientes consideraciones:

Figura 9. Descomposición de las variables



Fuente: Autor

- **Nombre de la variable:** Nombre de la variable identificada.
- **Descripción de la variable:** Se da una descripción detallada, de acuerdo con lo que hace referencia la variable.
- **Importancia de la variable:** Que tan sensible es la variable para hacer utilizada en el análisis.
- **Aporte de la variable:** Que tipo de aporte puede dar la variable en el análisis.
- **Indicador de la variable:** Fuente de información de la variable.
- **Periodo de tiempo:** Se define el periodo de tiempo de referencia a utilizar, si es en minutos, horas, días, semanas, meses o años.
- **Porcentaje de importancia:** De acuerdo con el aporte de la variable se define un porcentaje.

3.1.3. CARACTERIZACIÓN DE CADA VARIABLE

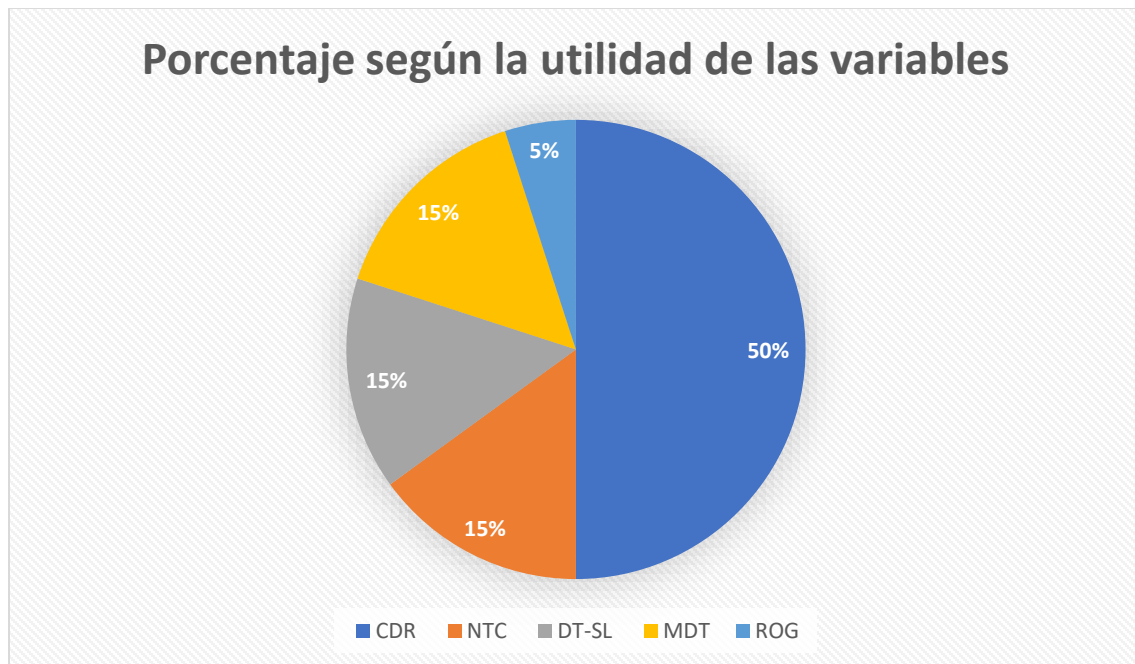
Tabla 2. Descripción detallada del proceso de caracterización de la variable.

| Nombre de la variable | Descripción de la variable | Importancia de la variable | Aporte de la variable | Indicador de la variable | Unidad de la variable | % |
|---------------------------------------|--|---|---|--|--|----|
| CDR-Registros de detalles de llamadas | Información de un registro de una llamada celular, de un proveedor de servicio celular. | Permite conocer los registros de llamada celular que se vayan a ver comprometidas con el análisis. | Registros de llamada celular. | La fuente de información de la variable sería las bases de datos de la radio bases en los que contengan todos los registros de la llamada celular. | El tiempo: Medido en periodos de tiempo cortos (10,15, 30 min.) de acuerdo con la zona en la que se esté realizando. | 50 |
| NTU-Numero de torres usadas | Cuenta el número de torres celulares, de las cuales una persona llamó en el período de tiempo. | Permite conocer las torres que se ven intervenidas cuando se realiza una llamada celular. | Cantidad de torres utilizadas para realizar el análisis | La zona y las torres identificadas, en la que se va a realizar el análisis de movilidad. | No aplica. | 15 |
| DT-SL-Distancia recorrida-línea recta | Es la suma de las distancias de línea recta, entre torres de las que se hicieron llamadas. | Permite conocer la distancia total entre todas las torres en la que se ven intervenida en el proceso de la llamada. | Saber la distancia de todo el recorrido de la persona que realizó la llamada. | La zona y las torres identificadas, en la que se va a realizar el análisis de movilidad. | No aplica. | 15 |
| MDT-Distancia máxima recorrida | Calcula la distancia máxima de línea recta entre dos torres que | Permite conocer la distancia recorrida de una torre a otra, que un usuario ha | Saber cuál es la próxima torre que un usuario utiliza en el proceso de | La zona y las torres identificadas, en la que se va a realizar el | No aplica. | 15 |

| | | | | | | |
|-------------------|---|---|--|--|------------|----|
| | una persona usó. | utilizado en una llamada. | una llamada | análisis de movilidad. | | |
| ROG-Radio de giro | Se determina, determinando primero el centro de masa de todas las torres celulares que una persona utiliza. | Permite conocer la distancia de centro de masa de la zona analizar. | Saber la distancia de ROG en una zona en particular. | Distancia total de todas las torres que utilizo en el proceso de la llamada. | No aplica. | 10 |

Fuente: Autor

Figura 10. Porcentaje según la utilización de las variables



Fuente: Autor

A partir de estas variables en el siguiente capítulo se propondrá un modelo de integración de cada una de las variables para la realización de una guía de análisis de movilidad en una zona específica.

En base a los CDR se derivan otras variables (NTC, DT-SL, MDT, ROG) que definen datos claves al proceso de movilidad, en los cuales son métodos analíticos que reflejan información fácil de interpretar.

Estas variables son tomadas de referencia de un análisis previo que se desarrolló en un artículo [10].

3.1.4. VARIABLES PARA MEJORAR EL ANÁLISIS DE MOVILIDAD

Las variables a continuación mencionadas son una mejora a las variables caracterizadas, que se subdividen en 3 grupos.

Grupo A. Medida de frecuencia de movilidad, número de viajes (NT).

Grupo B. Medida del rango espacial.

Grupo C. Medida de rango espacial y la frecuencia de la movilidad.

Se caracterizan por:

- Ser independientes de la densidad de la torre móvil y de las características sociales del contexto que influyen en la densidad de las torres.
- Son menos dependientes de la frecuencia de llamada de los usuarios, del movimiento en áreas sin cobertura de torre y de la dinámica temporal de la red celular subyacente de torres.
- Medir claramente los aspectos definidos de la movilidad, tales como la frecuencia y el alcance espacial del movimiento.
- Son relevantes y comparables entre contextos, países y tiempo.

Grupo A: Medida de frecuencia de movilidad, número de viajes (NT).

Esta medida es un recuento del número de veces que una persona realiza una llamada desde una celda de cuadrícula diferente a la llamada anterior. Si una persona hace una llamada desde una celda de la cuadrícula y su siguiente llamada es de la misma celda de la cuadrícula (independientemente de si es de una torre diferente) entonces este no es un viaje.

Grupo B: Medidas de la amplitud espacial de la movilidad.

Las dos medidas siguientes representan el número de lugares visitados por una persona. Al igual que con los viajes, una cuidadosa definición de lo que constituye un lugar se requiere para la coherencia y la comparabilidad a través de los contextos geográficos y el tiempo. Ambas medidas del grupo B requieren un supuesto de que todos los lugares en los que una persona existe por cualquier cantidad de tiempo podrían ser importantes. Algunos de estos lugares están marcados por una persona que hace una llamada. Sin embargo, hay otros lugares que una persona viaja a través de una ruta en la que no

hicieron una llamada. La lógica detrás de esta suposición es fundamentalmente la de un problema de datos faltantes: no se sabe cuánto tiempo una persona permanece en cada lugar, cuán importante era cada lugar a una persona en particular, o si los lugares donde hacían llamadas eran más o menos importantes que otros lugares por los que viajaban. En consecuencia, estas medidas suponen que todos los lugares a lo largo de la ruta de una persona de camino son de igual importancia y los cuenta todos.

Grupo C: Medidas de rango espacial y frecuencia de movilidad:

Las tres medidas de movilidad calculan la suma de las distancias entre los sitios donde ocurrieron episodios consecutivos de comunicación. Dividen sólo en términos del tipo de unidades de distancia utilizadas.

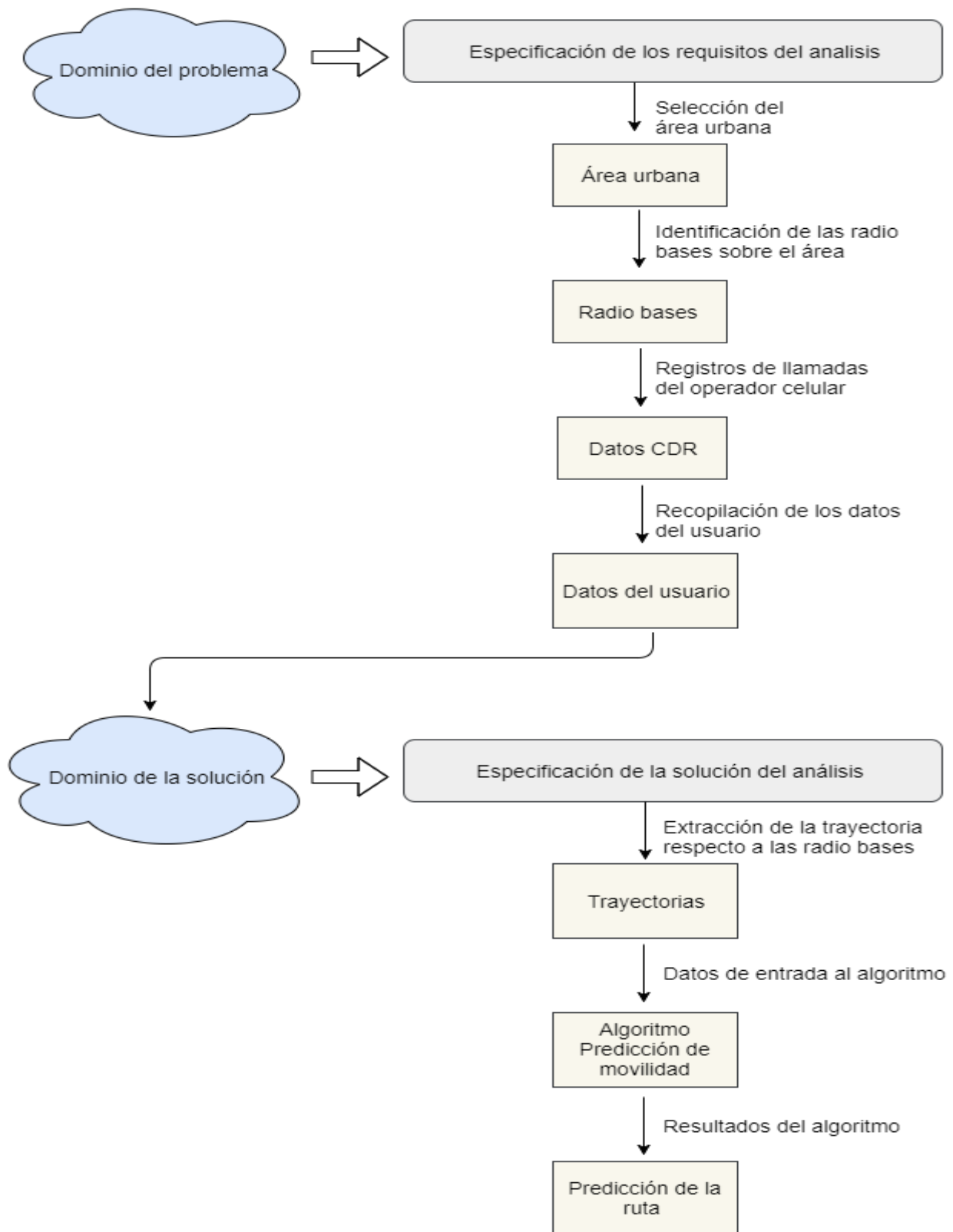
Estas distancias están relacionadas con la más rápida de todas las rutas posibles entre dos sitios que son consecutivos en una trayectoria espaciotemporal.

4. ANÁLISIS DEL PROCESO DE MOVILIDAD

En este capítulo se establecerá el proceso de unificar las variables caracterizadas en el capítulo anterior, que servirán de apoyo para realizar la guía del análisis de movilidad, partiendo de diferentes escenarios que se vayan a utilizar, y que permita llegar a un resultado o una conclusión sobre el estudio que se está realizando.

4.1. ESTRUCTURA DEL ANÁLISIS

Figura 11. Descripción del análisis de movilidad



Fuente: Autor

La estructura del análisis se subdivide en dos partes, en la especificación de los requisitos del análisis y en la especificación de la solución del análisis.

Especificación de los requisitos del análisis:

- **Área urbana:**

Parte de la especificación de los requisitos, en el cual primero se selecciona el área urbana en el que se realizara el análisis, esta área es una zona que haya un gran flujo de movilidad, y que se adapte para realizar el estudio.

- **Torre celular:**

Se identifican todas las torres celulares que se encuentren alrededor sobre el área de interés, y se seleccionan las torres celulares de los operadores de telefonía celular que va a suministrar los registros de llamadas.

- **Datos CDR:**

Los datos CDR serán los registros de llamadas que proporcione el operador de telefonía celular para realizar el análisis.

- **Datos del usuario:**

Son todos los CDR que tienen que ver directamente con el usuario que está involucrado con el análisis, aquí se recopilan los datos del usuario que quedan registrados en una torre celular cuando hace uso del servicio.

Especificación de la solución del análisis:

- **Trayectorias:**

Después de haber recopilado todos los datos, se extraen las trayectorias del usuario analizando cuáles fueron las torres celulares que registraron una conexión.

- **Algoritmo predicción de movilidad:**

Las trayectorias son los datos de entrada para el algoritmo de predicción de movilidad que se vaya a utilizar.

- **Predicción de la ruta:**

El algoritmo hará sus respectivos cálculos, generando al final una predicción de la ruta y comportamiento que se dio con relación a los registros de llamadas (CDR).

4.2. ESCENARIO BASE PARA EL ANÁLISIS DE MOVILIDAD

- Una persona que viaja sobre la avenida 30 agosto ubicada en la carrera 13 de Pereira, exactamente cerca al Parque Olaya Herrera, recibe una llamada mientras va conduciendo en su automóvil. ¿Cuál será el posible trayecto de esa persona?

Figura 12. Mapa de trayectoria



Fuente: Autor

4.2.1. CLASIFICACIÓN DE LA MOVILIDAD

La predicción de la movilidad puede clasificarse en dos grupos:

Basado en la historia: Se predice la siguiente celda o ruta del usuario basándose en las estadísticas de movilidad del usuario. Se registra el historial de movilidad del usuario y se obtiene la probabilidad de transición del usuario a la siguiente celda. Los métodos comunes para derivar la probabilidad de transición a la siguiente célula implican el modelo de cadena de Márkov, el modelo de Márkov oculto, las redes neuronales y el aprendizaje automático, la agrupación de rutas, etc.

Basada en la medición: Estos esquemas no se basan en el historial de movilidad del usuario, sino que derivan la probabilidad de la transición del usuario a la siguiente célula basada en mediciones en tiempo real (por ejemplo, RSSI, geometría, ángulo del usuario, distancia, etc.) hacen uso de la fuerza de la señal (RSSI, geometría) en dB (Geometría del usuario) para predecir la siguiente celda, donde se basa en el ángulo del usuario y las distancias para predecir la siguiente celda.

Estos esquemas de predicción de la movilidad se utilizan para reservar proactivamente los recursos, activar el balanceo de carga (LB) o activar / desactivar las celdas pequeñas. Por lo tanto, la predicción de la movilidad puede verse como una fuerza motriz para la conciencia del contexto en las redes celulares.

4.2.2. ANÁLISIS DE MOVILIDAD EN BASE A LAS CADENAS DE MÁRKOV

Para el análisis de movilidad se tendrá en cuenta el algoritmo explicado en el capítulo 2, marco para apoyar el conocimiento del contexto de movilidad en redes celulares, que hace referencia a las cadenas de Márkov. El análisis del algoritmo requiere de la siguiente información.

La posición del usuario: Esta información puede ser en forma de coordenadas (x, y, z) y podría ser obtenida mediante sistemas de posicionamiento global (GPS).

La velocidad de usuario: Esta información puede obtenerse a través de mediciones Doppler o se puede obtener de indicador de velocidad de vehículo mediante el uso de la interconexión apropiada de infraestructura vehicular a la red celular.

La geometría de usuario (dB): Medida en el terminal de usuario.

Lista de celdas vecinas (NCL): Mantenido por el sistema de soporte de operaciones (OSS).

Mapas de rutas: Similar a NCL, pero tiene información sobre las carreteras, caminos cruzados, los posibles agujeros de cobertura en ellos.

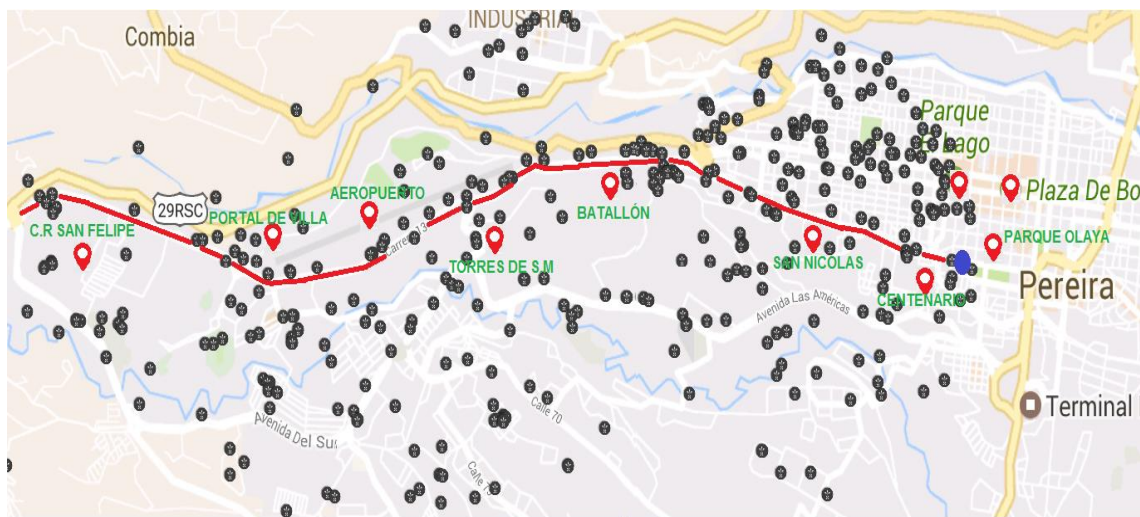
Información infraestructura vehicular

Se propone para extraer y explotar la información siguiente de la infraestructura vehicular para mejorar la predicción de la movilidad y la sensibilidad al contexto.

Origen: La posición inicial (ubicación), desde donde el usuario inicia su viaje.

Destino: La posición final (ubicación) a donde el usuario tiene la intención de viajar

Figura 13. Mapa de la trayectoria con posibles lugares a dirigirse



Fuente: Autor

Se considera la vista del escenario de simulación que se representa en la figura 13, se hace referencia a Pereira. Hay 10 puntos de referencia que pueden actuar como origen y destino del usuario (Parque Olaya, Plaza de Bolívar, Parque el lago, B. Centenario, B. San Nicolás, Batallón, Torres de San Mateo, Aeropuerto, Portal de la Villa, Conjunto Residencial San Felipe). La probabilidad de transición de un usuario desde la estación

base BS_n a BS_{n+1} basado en predicción de la siguiente celda en función de Márkov se puede obtener por,

$$P(BS_n \rightarrow BS_{n+1}) = \frac{N(BS_n \rightarrow BS_{n+1})}{N(BS_n)} \quad (\text{Ec. 1, Cap 4.2.2})$$

Dónde, $BS_n \rightarrow BS_{n+1}$ indica transición de $Cell_n$ a $Cell_{n+1}$ n, indica el número de veces. Por lo tanto, mediante el uso de estadísticas de movilidad, la probabilidad de transición en una celda siguiente se puede obtener como la proporción entre; número de veces que el usuario de $Cell_n$ transitado a $Cell_{n+1}$ y el número de veces que el usuario se encuentra en $Cell_n$.

Este modelo se puede mejorar mediante el uso de información de contexto adicional, como origen del usuario (O) y su destino (D). Las ecuaciones de probabilidad resultantes son:

$$P(BS_n \rightarrow BS_{n+1}/O) = \frac{N(BS_n \rightarrow BS_{n+1}/O)}{N(BS_n/O)} \quad (\text{Ec. 2, Cap 4.2.2})$$

$$P(BS_n \rightarrow BS_{n+1}/O\&D) = \frac{N(BS_n \rightarrow BS_{n+1}/O\&D)}{N(BS_n/O\&D)} \quad (\text{Ec. 3, Cap 4.2.2})$$

En la ecuación 2, se consideran las estadísticas del usuario sólo cuando la ruta comienza a partir de un origen en particular, sin considerar su destino, mientras que en la ec. 3 se consideran las estadísticas de usuario cuando la ruta se origina en un lugar específico y viajan a un destino específico.

Del mismo modo, el siguiente trayecto de una ruta (carretera) de un usuario se puede predecir antes de que alcance el cruce. Esto también es análogo a la predicción de la probabilidad a su vez en cada cruce. La probabilidad de transición de un usuario de la carretera R_n a R_{n+1} con base en la predicción de Márkov, estaría dada por la ecuación,

$$P(R_n \rightarrow R_{n+1}) = \frac{N(R_n \rightarrow R_{n+1})}{N(R_n)} \quad (\text{Ec. 4, Cap 4.2.2})$$

El modelo de Márkov puede mejorarse aún más mediante la explotación de la información de contexto sobre el origen del usuario (O) y su destino (D). Las probabilidades respectivas se dan como,

$$P(R_n \rightarrow R_{n+1}/O) = \frac{N(R_n \rightarrow R_{n+1}/O)}{N(R_n/O)} \quad (\text{Ec. 5, Cap 4.2.2})$$

$$P(R_n \rightarrow R_{n+1}/O\&D) = \frac{N(R_n \rightarrow R_{n+1}/O\&D)}{N(R_n/O\&D)} \quad (\text{Ec. 6, Cap 4.2.2})$$

En ec. 5, se consideran las estadísticas de usuario sólo cuando la ruta se originó a partir de un origen determinado, en tanto que, en la ec. 6 se consideran las estadísticas del usuario cuando la ruta se originó a partir de un origen específico y que viaja a un destino específico para predecir la siguiente carretera (ruta) del usuario.

Simulación

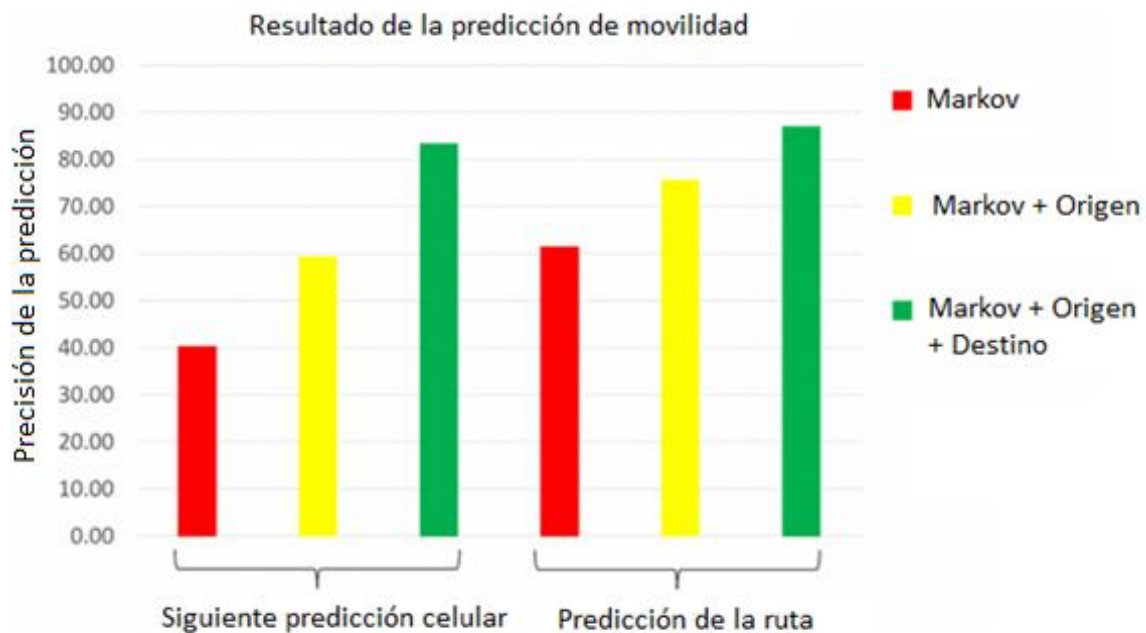
A continuación, se presenta una simulación tomada del artículo [6], podría ser un tipo de análisis que se requiere en una situación real y que se presenta para mayor claridad sobre el proceso a seguir.

Un automóvil que viaja a 30 km / h está presente y atraviesa un conjunto de 10 trayectorias, cada una de las cuales incluye los 10 puntos de referencia. Se supone que el coche sigue una movilidad diurna, donde atraviesa a lo largo de un conjunto de trayectorias fijas diariamente en lugar de seguir la movilidad aleatoria de puntos de paso. Las estaciones base se habilitan con tecnología LTE y parámetros de simulación (por ejemplo, parámetros de antena, modelos de pérdida de ruta, etc.) siguen las pautas UIT-R, METIS. El ancho de banda es de 10 MHz (50 PRB) y la frecuencia portadora es de 2 GHz. Se realizan 100 ejecuciones de simulación para obtener estadísticas de movilidad del usuario. A continuación, se realizan los esquemas de predicción de movilidad

anteriormente mencionados para otras 100 simulaciones y se calcula la precisión media de la predicción. La comparación entre los diferentes esquemas se muestra en la figura 14. Se pudo observar que la precisión de predicción aumenta cuando se utiliza información de contexto adicional sobre el origen y el destino del usuario. La precisión de la predicción de la celda siguiente se mejora a alrededor del 85% y la precisión de predicción de la ruta se mejora a alrededor del 90%, lo que indica que extraer y explotar información sobre el origen y el destino del usuario de la infraestructura vehicular es valiosa para la predicción de la movilidad.

Los resultados de la predicción de movilidad son los siguientes:

Figura 14. Predicción de la movilidad



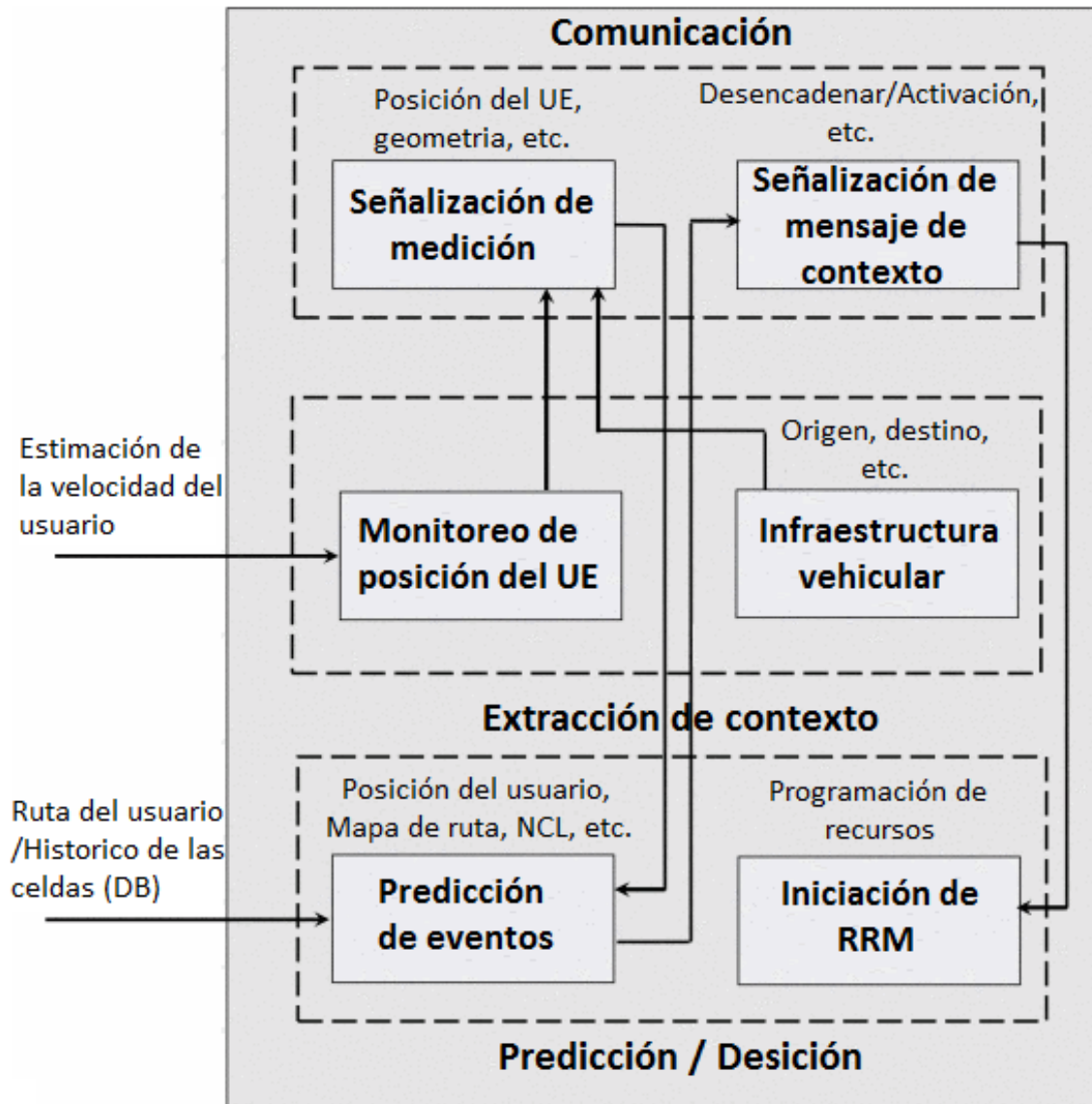
Fuente: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/7504249/>.

Con base a estos resultados del estudio de movilidad, deja claro que el algoritmo de Márkov es un gran aporte para mejorar el estudio de análisis de movilidad, teniendo en cuenta la información del usuario como: La posición del usuario, la velocidad de usuario, la lista de celdas vecinas, y el mapa de rutas.

Interfaces para apoyar el conocimiento del contexto de movilidad

Los procedimientos de contexto de movilidad pueden ser descompuestos funcionalmente en tres bloques como se muestra en la figura 15.

Figura 15. Descomposición funcional



Fuente: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/7504249/>.

Comunicación: Este bloque trata de señalar información entre el usuario y la estación base. Las dos partes que lo componen son:

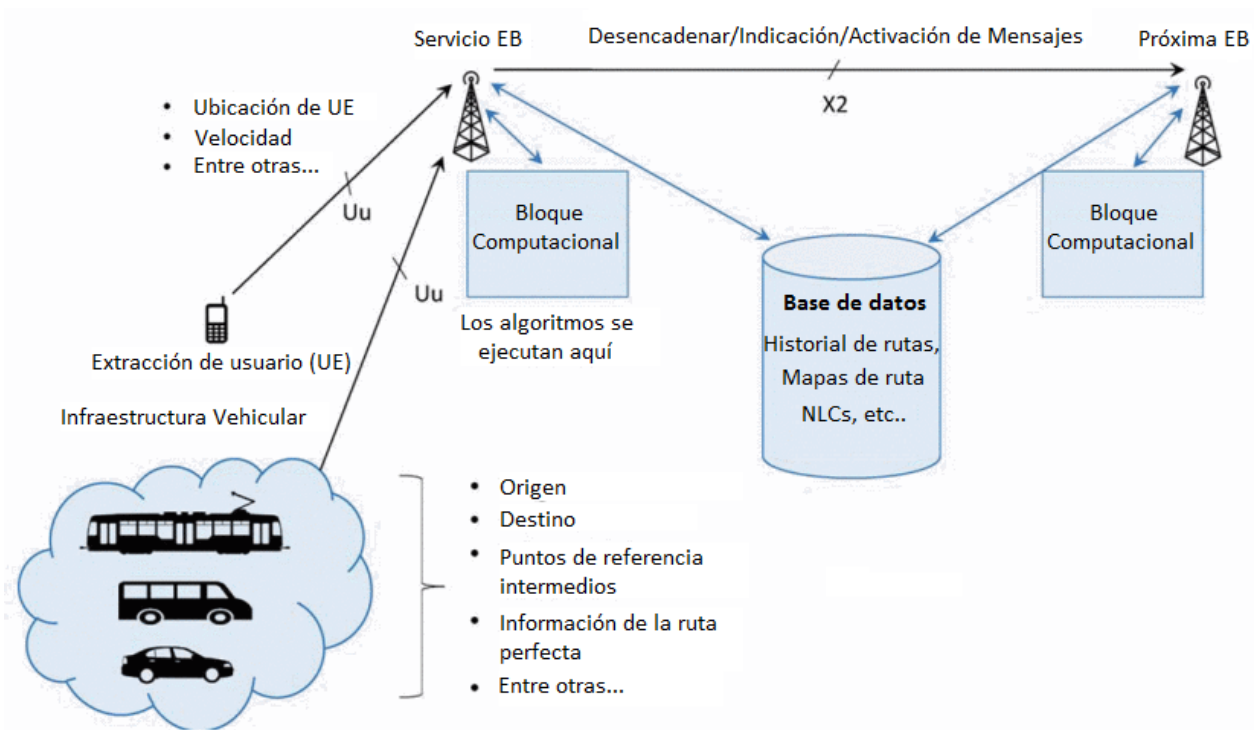
a) Señalización de medición, donde la conforman la posición de UE (Extracción del usuario), geometría, etc.

b) Mensaje de contexto de señalización (por ejemplo, disparador para equilibrar la carga / gestión de recurso, mensaje de activación / desactivación de las celdas, etc.) desde la estación base de servicio a la estación base de destino (predicha).

Extracción de contexto: Se genera a partir de la interacción de dos variables, monitoreo de posición de UE y de la infraestructura vehicular, con base a estas dos variables es posible obtener posiciones de usuario, medición de la intensidad de la señal / geometría, obtener información sobre el origen y destino de la infraestructura vehicular, etc., son incluidas en este bloque. Adicionalmente la estimación de la velocidad se requiere para establecer la tasa del muestreo de la información antes mencionada.

Predicción / Decisión: La conforman la predicción de eventos, que es la posición del usuario, mapa de rutas y NCL (Lista de celdas vecinas), la iniciación de RMM (Medición de rutas de recursos) que es la programación de recursos, tienen lugar en este bloque, y la gestión de recursos, las decisiones de activación / desactivación de celdas, etc., también se hacen aquí.

Figura 16 Arquitectura General



Fuente: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/7504249/>.

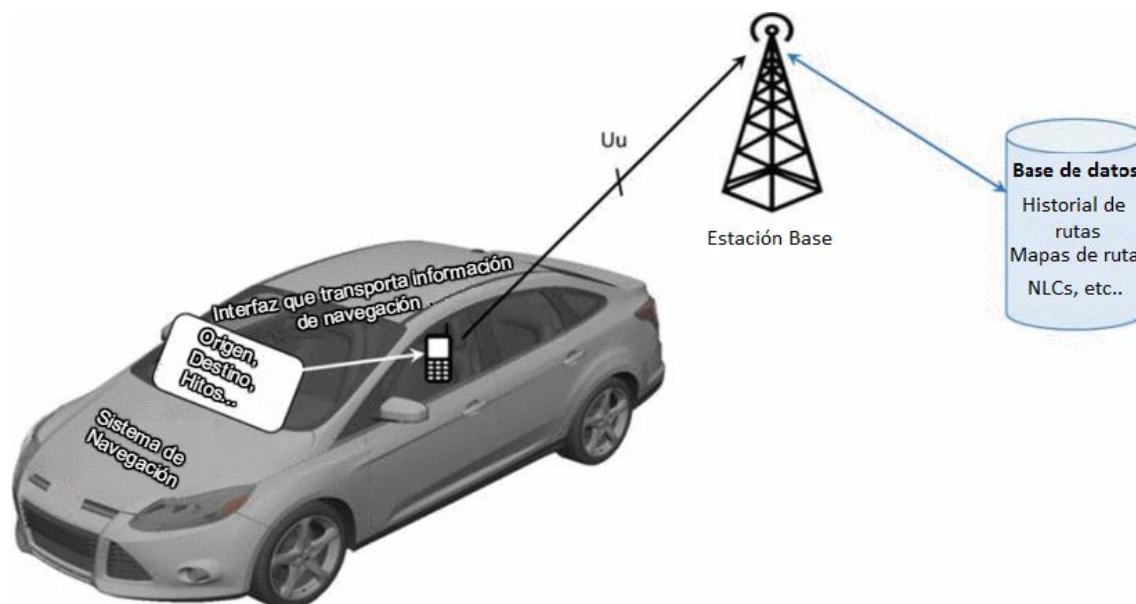
Basado en el análisis de descomposición funcional, la arquitectura general para apoyar la conciencia del contexto de movilidad se muestra en la figura 16.

La información sobre las posiciones del usuario, la geometría (dB), la velocidad, etc., deben ser señaladas a la estación base de servicio a través de la interfaz Uu. Además, se requiere que la información proveniente de infraestructura vehicular como origen, destino, puntos de referencia intermedios, información de ruta perfecta (que puede ser factible en un futuro próximo debido a avances en automóviles de conducción autónomos) sea comunicada a la estación base de servicio a través de la interfaz Uu. Se requiere una base de datos para almacenar el historial de la movilidad del usuario, la lista de celdas vecinas, los mapas de rutas de las respectivas regiones, y así sucesivamente. Se requiere una interfaz adecuada para pasar la información de la movilidad del usuario a la base de datos. Las estadísticas de movilidad de un usuario se almacenan en la base

de datos a través de la estación base de servicio. Los bloques computacionales en la estación base de servicio son responsables de la predicción (por ejemplo, la siguiente celda) adquiriendo información del usuario y haciendo interfaz con la base de datos. El mensaje de contexto subsiguiente (por ejemplo, activación de LB, activación de celdas) necesita ser transportado a la célula objetivo a través de la interfaz X2

Es evidente que la información de contexto sobre el origen y el destino del usuario son valiosa para mejorar la conciencia del contexto de movilidad. Con el fin de extraer y explotar eficientemente dicha información de la infraestructura vehicular, se requieren nuevas interfaces. La Figura 17 describe la nueva interfaz prevista entre el sistema de navegación del vehículo y el equipo de usuario, para adquirir información sobre origen, destino, información de ruta, etc. y la señala además a la estación base de servicio a través de la interfaz Uu.

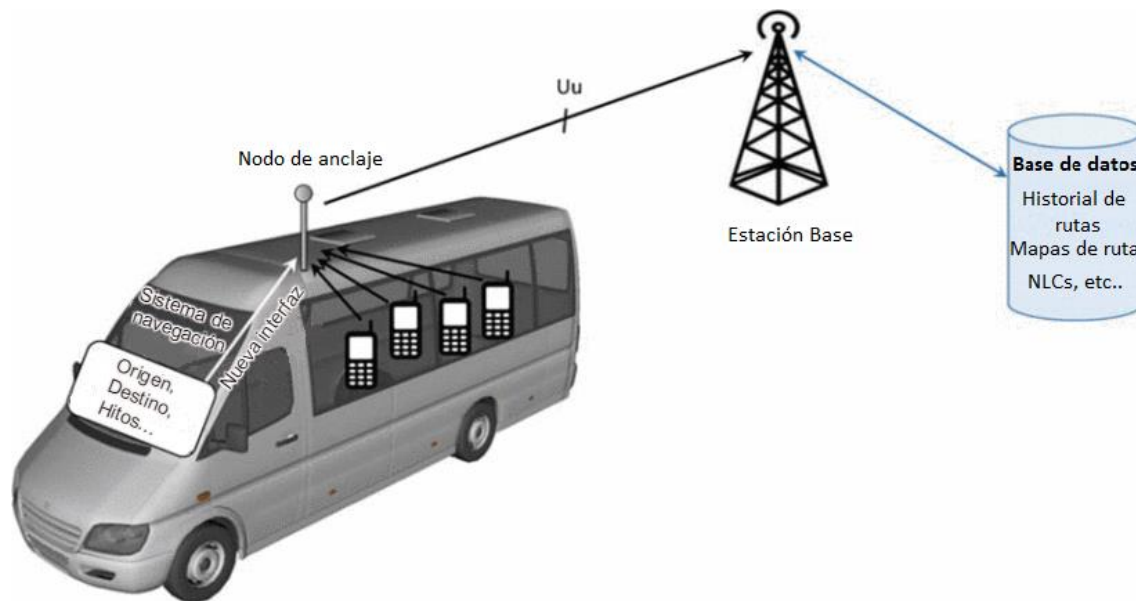
Figura 17. Interfaz en infraestructura vehicular



Fuente: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/7504249/>.

En el caso de varios usuarios que viajan juntos (por ejemplo, transporte público como autobús, etc.), el número de interfaces y señalización puede minimizarse al tener un nodo de anclaje controlando a los usuarios en el vehículo, formando una red en movimiento. En tal caso se requiere una nueva interfaz entre el nodo de anclaje y el sistema de navegación como se describe en la figura 18.

Figura 18. Interfaz en las redes en movimiento



Fuente: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/7504249/>.

4.3. ANÁLISIS DE MOVILIDAD EN BASE A LA MEJORA DE LAS VARIABLES CARACTERIZADAS

Para el análisis con las variables caracterizadas se toma de referencia la avenida 30 agosto donde una persona (P), realiza dos llamadas desde dos sitios diferentes, la primera de una antena de telefonía móvil situada con la etiqueta "sitio 766" y la segunda de una antena de telefonía móvil situada con la etiqueta "sitio 899", en total hay 1536 cuadrículas celular indexadas de 0.5 km x 0.5 km entre 1 (Que es la celda en la esquina izquierda inferior) y 1536 (la celda de la esquina superior derecha).

Un sitio es una celda de cuadrícula que contiene al menos una torre celular, en el mapa se muestra la trayectoria recta entre los sitios de 776 y 899, y también la ubicación de la ruta de carretera más rápido que es la ruta con el tiempo de viaje estimado más pequeño entre los dos sitios.

Figura 19. Mapa de la trayectoria de avenida 30 de agosto, referenciado con cuadrículas



Radio bases activas



Avenida 30 de agosto



Llamada 1 Sitio 766



Llamada 2 Sitio 899

Fuente: Autor

Comparar esta persona a una segunda persona que vive en una zona rural con una sola torre en un radio de 5 kilómetros alrededor de su casa. Incluso sin ellos moverse alrededor del radio de 5 kilómetros tan a menudo como la persona urbana, rural, el individuo podría utilizar únicamente ésta torre y así ser clasificados como no se mueve en cualquier lugar y alcanzar la calificación más baja movilidad. Por lo tanto, si no se toma en cuenta, las variaciones en la densidad de la torre crean variaciones en la movilidad

que en realidad no existen. Este problema se agrava aún más por el hecho de que las torres celulares se colocan con mayor frecuencia en las zonas urbanas con alta densidad de población, zonas políticamente importantes, tales como ciudades capitales o regiones más favorecidas con mayor penetración de telefonía móvil.

Una segunda preocupación, es que la colocación de torres celulares varía con el tiempo. En muchos países, donde la infraestructura móvil aún no ha alcanzado la saturación, Las nuevas torres celulares se construyen cada año para dar cabida a un número creciente de usuarios.

Un tercer problema es que las medidas existentes se confunden con la densidad de la torre, también son inherentemente confundidos con la frecuencia de llamada. Cuanto más a menudo una persona llama, se registrarán en más torres. Una persona que usa su teléfono con frecuencia es probable que tenga una calificación de movilidad diferente, en comparación con una persona con la misma trayectoria espaciotemporal, pero a menor frecuencia de llamada.

Un cuarto problema es que las medidas existentes de movilidad se basan fundamentalmente en suposiciones implícitas, sin embargo, poco realistas acerca de la naturaleza del movimiento humano. Sus definiciones implican la medición de distancias en líneas rectas entre las torres celulares. De hecho, los seres humanos casi nunca viajan en línea recta y fuera del transporte aéreo que no suelen aparecer en un solo lugar, luego desaparecen y aparecen de nuevo en otro lugar distante.

El quinto problema es que no está del todo claro qué aspecto de movilidad estas medidas captura. Debido a la densidad variable de torres celulares, la medida NTU no captura las regiones espaciales. Sin embargo, debido a que cuenta con torres únicas, tampoco evalúa la frecuencia de movimiento. La medida DT-SL capta tanto la frecuencia de movimiento como el rango espacial. La medida MDT, ya que incorpora sólo dos de las torres utilizadas, no capta la frecuencia ni de regiones espaciales bien. La medida ROG no captura la frecuencia de movimiento. A pesar de que inicialmente parece capturar rango espacial, lo hace de una manera inconsistente que es influenciada por la frecuencia de llamada de cada torre utilizada.

Teniendo en cuenta estas preocupaciones acerca de las medidas existentes de movilidad, se plantean nuevas medidas que son:

- (i) Son independientes de la densidad de la torre móvil y las características sociales de contexto que influyen en la densidad de la torre.
- (ii) Son menos dependientes de la frecuencia de llamadas de los usuarios, el movimiento en áreas sin cobertura de la torre, y la dinámica temporal de la red celular subyacente de las torres.
- (iii) Medir los aspectos claramente definidos de la movilidad, tales como la frecuencia y el alcance espacial de movimiento.
- (iv) Son relevantes y comparables entre los contextos, los países, y el tiempo.

La primera base de las medidas es un sistema de celdas de cuadrícula de tamaño uniforme colocado a través de una ciudad o área de estudio. Algunas celdas de la rejilla tienen una torre celular en ellas, otras no y algunas tienen múltiples torres celulares. Se hace referencia a una celda de rejilla con al menos una torre activa como un sitio. Con el sistema de cuadrícula, si por ejemplo de una persona, R, llama desde una torre celular, se registran como localizados en el centroide del sitio correspondiente (celda de cuadrícula). El movimiento se calcula sólo cuando R se mueve de un sitio a otro. Si R llama de nuevo desde otra torre en el mismo sitio, entonces están registrados en el mismo sitio, y por lo tanto no se han movido. Pero si la próxima llamada que R hace es manejada por una torre en un sitio diferente, entonces se han movido. Al reemplazar las torres celulares con sitios, el efecto adverso de la variabilidad temporal de la extensión espacial de la cobertura de las torres celulares también está significativamente disminuido. En un período de tiempo dado, una torre es activa si maneja al menos una comunicación celular durante ese período. De lo contrario, una torre está inactiva y no contribuye a la creación de un sitio.

El segundo fundamento de las medidas es un conjunto de suposiciones realistas acerca de cómo viajan los seres humanos: generalmente utilizan carreteras, toman la ruta más rápida y accesible de un lugar a otro y la velocidad de los viajes se ve afectada por los

límites de velocidad y la calidad de superficies de carreteras. Con estas suposiciones, se utilizan los datos de SIG disponibles públicamente sobre sistemas de carreteras para crear rutas de viaje de un lugar a otro que no sean líneas rectas. Con esta información, es posible calcular una ruta de viaje asumida entre dos puntos cualesquiera de una zona, donde la ruta asumida tiene el tiempo de viaje más corto posible comparado con todas las demás rutas. Debido a que todas las medidas se basan en un sistema de cuadrícula, se utilizan rutas de viaje asumidas que comienzan en el centroide de un sitio desde el que una persona hizo una llamada.

La tercera base de las medidas es que los seres humanos viajan con más frecuencia en el suelo. Incluso si no hacen llamadas en todos los lugares que visitan, se puede suponer que existieron durante algún tiempo en cada lugar a lo largo de una ruta de carretera, entre dos llamadas posteriores. Esta suposición mejora parcialmente la influencia de confusión de la frecuencia de llamada y no hay torres disponibles en la medición de la movilidad.

En base a estos fundamentos se crean seis nuevas medidas de movilidad, y se dividen en tres grupos dependiendo de cuál de las dos dimensiones clave de la movilidad que capturan.

El *Grupo A* incluye medidas que captan la frecuencia de la movilidad, pero no captan rango espacial.

El *Grupo B* incluye medidas que capturan el rango espacial, pero no la frecuencia.

El *Grupo C* incluye medidas que captan tanto la frecuencia y el alcance espacial.

Las medidas dentro de los grupos se relacionan, y por lo tanto se correlacionan fuertemente. A continuación, se describen estas nuevas medidas sus beneficios y sus limitaciones.

Las combinaciones de medidas que pertenecen a cada uno de estos tres grupos son necesarias para identificar diversos patrones de movilidad que existen en una población o una zona. Por ejemplo, una persona P realiza un solo viaje largo, por lo que su movilidad será clasificada como alta por las medidas del grupo B, pero no por las medidas en los

grupos A y C. Considerar otras dos personas que llaman P1 y P2 que van de casa a trabajo y de vuelta por 20 días cada mes, pero el trabajo de P1 es 1 km de su casa y el trabajo de P2 es de 10 km de su casa. Por lo tanto, se mueven con la misma frecuencia en un período de tiempo dado, pero las distancias entre lugares consecutivos en la trayectoria de P1 son más cortos que los de la trayectoria de P2. En este caso, la movilidad de P1 y P2 será igual cuando se evalúe por medidas del grupo A, pero se diferenciará cuando se evalúe por medidas de los grupos B y C. La medida o combinación de medidas que se use dependerá del contexto de la investigación que se quiera analizar.

Grupo A: Medidas de frecuencia de movilidad. Número de viajes (NT). Esta medida es un recuento del número de veces que una persona realiza una llamada desde una celda de cuadrícula diferente a la llamada anterior. Si una persona hace una llamada desde una celda de la cuadrícula y su siguiente llamada es de la misma celda de la cuadrícula (independientemente de si es de una torre diferente) entonces este no es un viaje. P realiza dos llamadas desde dos sitios diferentes, por lo que el valor de NT es igual a 1 (percentil 10). Si P hubiera realizado cualquier número de llamadas subsiguientes utilizando sólo las dos torres del sitio 766, el valor de NT permanecería sin cambios. Esta medida no depende de hasta qué punto una persona viaja (rango espacial): P podría haber llamado desde cualquiera de los dos sitios activos, y el valor de NT sería el mismo. Pero si P hace una llamada desde otro sitio, el valor de NT aumentará en 1. La cantidad de tiempo entre las llamadas no se tiene en cuenta al calcular NT.

Las limitaciones de NT provienen de la información incompleta sobre movilidad contenida en CDRs. Esta medida se basa en una especificación de un viaje como un movimiento entre dos lugares donde una persona existió por cualquier cantidad de tiempo. La literatura de transporte a menudo define un viaje como un movimiento entre dos lugares donde una persona permaneció por un período mínimo de tiempo (a menudo 5 o 10 minutos). Usando CDRs, no es posible determinar cuánto tiempo una persona se quedó en cada lugar que hizo una llamada telefónica. Esta medida derivada de la CDR podría registrar menos viajes si se requiere un tiempo mínimo en un destino o si una persona no hace una llamada cuando en un destino particular antes de salir para su próximo destino.

Grupo B: Medidas de la amplitud espacial de la movilidad. Las dos medidas siguientes representan el número de lugares visitados por una persona. Al igual que con los viajes, se requiere una definición cuidadosa de lo que constituye un lugar para la consistencia y comparabilidad a través de los contextos geográficos y el tiempo. Ambas medidas del grupo B requieren un supuesto de que todos los lugares en los que una persona existe por cualquier cantidad de tiempo podrían ser importante. Algunos de estos lugares están marcados por una persona que hace una llamada. Sin embargo, hay otros lugares que una persona viaja a través de una ruta en la que no hicieron una llamada. La lógica detrás de esta suposición es fundamentalmente la de un problema de datos que falta: no se sabe cuánto tiempo una persona permanece en cada lugar, cuán importante era cada lugar para una persona en particular, o si los lugares donde hicieron llamadas eran más o menos importantes que otros lugares por los que viajaban. En consecuencia, estas medidas suponen que todos los lugares a lo largo de la ruta de una persona de camino son de igual importancia y los cuenta todos.

Para calcular las medidas del grupo B, se toman cada par de sitios que son consecutivos en una trayectoria espaciotemporal M y se identifica las celdas de cuadrícula que pertenecen a la ruta más rápida entre los dos sitios. Se forma el conjunto de todas las celdas $V(M)$ de la rejilla en estas rutas de carretera más rápidas que también incluyen sus celdas de inicio y de fin, los sitios desde los que se realizaron las llamadas. Una celda de cuadrícula aparece sólo una vez en $V(M)$. Los elementos de $V(M)$ se denominan celdas de cuadrícula visitadas. Las celdas de cuadrícula visitadas que tienen torres celulares en ellas se llaman sitios visitados.

Las celdas de cuadrícula visitadas (GCV-R). Esta medida viene dada por el número de celdas de cuadrícula visitadas.

Sitios visitados (SV-R). Esta medida es la relación entre el número de sitios visitados y el número total de sitios en el período de referencia de la trayectoria M . El número de sitios podría cambiar de un período de referencia a otro como las torres celulares están instaladas o desmanteladas. Por tanto, es necesario ajustar para el número variable de sitios en el tiempo para definir una medida cuyos valores son consistentes a través de los períodos de tiempo de referencia. La definición de la medida SV-R se basa en la

suposición de que hay algo importante acerca de dónde se ubica una torre celular, ya sea alta densidad de población o de un área importante, la razón por la cual las torres celulares se encuentran en ciertas áreas puede variar entre los contextos a lo largo del tiempo.

Se usa esta suposición particular porque asume lo menos posible para definir un lugar y por lo tanto es el más comparable a través de contextos y tiempo. Estas dos medidas tienen en cuenta el rango espacial de la movilidad de una persona: cuanto más lejos cada viaje, mayor es el número de sitios y celdas visitadas. Pero la frecuencia de movimiento no es captada por estas medidas. Si P realizara una tercera llamada desde el sitio 899, una cuarta llamada desde el sitio 766, una llamada del sitio 899 y así sucesivamente, los valores de GCV-R y SV-R permanecerán igual.

Grupo C: Medidas de rango espacial y frecuencia de movilidad. Las tres medidas de movilidad calculan la suma de las distancias entre los sitios donde ocurrieron episodios consecutivos de comunicación. Dividen sólo en términos del tipo de unidades de distancia utilizadas. Estas distancias se relacionan con la más rápida de todas las rutas posibles entre dos sitios que son consecutivos en una trayectoria espaciotemporal.

Distancia recorrida (DT-R). La métrica de distancia para esta medida es la longitud de la ruta más rápida. Existen dos diferencias clave entre el DT-R y la medida existente DT-SL: (i) DT-SL implica movimiento entre torres celulares, mientras que DT-R implica movimiento entre sitios; Y (ii) DT-SL es la suma de las distancias de línea recta, mientras que DT-R es la suma de las distancias a través del recorrido por carretera. Si se realizan dos llamadas consecutivas usando dos torres que pertenecen al mismo sitio, DT-SL registrará la distancia entre las dos torres mientras que DT-R no registrará ningún movimiento. Por otro lado, DT-SL subestimaré las distancias entre dos puntos ya que las distancias de línea recta son casi siempre (si no siempre) más pequeñas que las distancias por carreteras.

Tiempo recorrido (TT-R). La métrica de distancia para esta medida es el tiempo de viaje estimado en la ruta más rápida. El tiempo de viaje puede estimarse de varias maneras. Los límites de velocidad pueden utilizarse cuando estén disponibles. Si no se dispone de

información de límite de velocidad o si la calidad de las carreteras es tal que no se pueden cumplir los límites de velocidad, se deben estimar las velocidades de desplazamiento promedio para cada tipo de carretera, ya que este es el menor tiempo de viaje estimado por carreteras entre los dos sitios desde los que se realizaron las llamadas.

Viaje de las celdas de cuadrícula (GCT-R). La métrica de distancia para esta medida es el número de celdas de cuadrícula interceptadas por la ruta más rápida. El sitio inicial de una ruta se cuenta, pero el sitio final no se cuenta. Los sitios que son el final de una ruta y el principio de otra no se cuentan dos veces.

Estas tres medidas incorporan la frecuencia y el rango espacial de la movilidad de una persona. Cuantas más veces una persona se mueva, mayor será la distancia, el tiempo y las celdas de la rejilla. Cuanto más lejos cada viaje, mayores serán estas medidas. Cuando se emplean conjuntamente, estas tres métricas de distancia son útiles para identificar diversos patrones de movilidad. Por ejemplo, la movilidad de dos individuos que viajan a la misma distancia, pero utilizan diferentes tipos de carreteras (por ejemplo, carreteras vs. caminos secundarios) se clasifica igual por DT-R, pero será diferente con respecto a TT-R. La métrica de distancia para el GCT-R depende menos de formas particulares de las carreteras o de su calidad, lo que podría ayudar a comparar la movilidad de trayectorias espaciales registradas en regiones o países distantes.

4.3.1. EVALUACIÓN DE LAS MEDIDAS PROPUESTAS DE MOVILIDAD

La evaluación de las medidas de movilidad de la CDR está limitada por la realidad de que en la actualidad no existe una medida estándar de movilidad o no hay un patrón oro al que se pueda comparar nuevas medidas. En este sentido, la herramienta de evaluación más importante disponible es la validez de cara. En otras palabras, la mejor herramienta de evaluación es una cuidadosa discusión de qué medidas tienen sentido y si realmente miden lo que se cree que deberían estar midiendo. Parte de esta discusión de validez de cara está arriba en la descripción de las medidas, dimensiones de movilidad, y suposiciones requeridas para cada medida.

Además de la validez aparente y correlaciones, es importante evaluar qué grupos de medidas y qué medidas dentro de cada grupo se debe utilizar para los estudios de

movilidad de la población. Las seis medidas que se introducen son múltiples opciones de combinaciones que podrían seleccionarse sobre la base de preguntas y contextos particulares de investigación. También se argumenta que las seis medidas son necesarias en los estudios de movilidad de la población de principio a fin. A pesar de tener características comunes en términos de las dos dimensiones clave de la movilidad que se han discutido, cada medida capta un aspecto ligeramente diferente de la movilidad, por lo que es relevante solo y junto con las otras cinco medidas.

La medida de movilidad del tiempo recorrido (TT-R) tiene un uso adicional importante, los valores de esta medida pueden utilizarse para identificar y posiblemente detectar las trayectorias espaciotemporales afectadas por errores en las bases de datos del proveedor de servicios celulares o por intrusos que obtienen acceso no autorizado a teléfonos móviles y los utilizan para comunicarse al mismo tiempo que los propietarios reales. La identificación de estas trayectorias inusuales no es posible con las medidas de movilidad existentes.

4.3.2. DEFINICIONES MATEMÁTICAS DE LAS MEDIDAS DE MOVILIDAD

Definiciones matemáticas formales de las medidas de movilidad descritas anteriormente. Se considere la secuencia de CDR asociada con un teléfono móvil en un período de tiempo de referencia T (por ejemplo, un día, una semana):

$$(Ec. 1, Cap. 4.3.2) \quad M = \{m_1, m_2, \dots, m_n\}$$

Suponiendo que el proveedor de servicios inalámbricos que generó los CDR tiene torres activas K en el período de tiempo de referencia T y que las ubicaciones espaciales $l_i^{CT}, i \in K = \{1, 2, \dots, K\}$ de estas torres activas son conocidas. En (1), $m_i \in K, 1 \leq i \leq n$, es el identificador de la torre celular que maneja la comunicación representada por la i -ésima CDR en la secuencia (1). Si $i < j$ la comunicación representada por m_i fue registrada antes de la comunicación representada por m_j . Se hace referencia a M como la trayectoria espaciotemporal del teléfono celular que generó la secuencia de CDRs y que más de una torre pudo haber manejado la misma comunicación (por ejemplo, una llamada), pero en este caso varios CDR - una para cada torre celular - puede haber sido generada.

Para cualquier par de localizaciones espaciales l y l' identificadas por sus coordenadas de latitud y longitud, se define la función de distancia $d_{SL}(l, l')$ que representa la distancia de línea recta entre l y l' . Se toma $d_{SL}(l, l) = 0$.

Medidas existentes de movilidad

Para la trayectoria espaciotemporal M de (1), la medida de movilidad llamada "número de torres utilizadas" (NTU) es el número de torres únicas que aparecen en esta secuencia, es decir.

$$\#\{i : i \in K \text{ Tal que existe } m_j, 1 \leq j \leq n \text{ con } m_j = i\}$$

Aquí $\#A$ denota el número de elementos en el conjunto A . La medida de movilidad llamada "distancia recorrida" (DT-SL) es la suma de las distancias entre torres consecutivas a partir de las cuales ocurrió la comunicación:

$$\sum_{j=2}^n d_{SL}(l_{m_{j-1}}^{CT}, l_{m_j}^{CT}).$$

La medida de movilidad llamada "distancia máxima recorrida" (MDT) es la distancia máxima en línea recta entre dos torres en la secuencia M :

$$\max_{1 \leq i < j \leq n} d_{SL}(l_{m_i}^{CT}, l_{m_j}^{CT}).$$

La medida de movilidad denominada "radio de giro" (RoG) es la raíz cuadrada de la media de las distancias rectas cuadradas entre las ubicaciones de las torres en M y el centro de masa.

$$\bar{l}_M = \frac{1}{n} \sum_{j=1}^n l_{m_j}^T, \quad (\text{Ec. 2, 4.3.2})$$

De la trayectoria:

$$\sqrt{\frac{1}{n} \sum_{j=1}^n d_{SL}^2(l_{m_j}^T, \bar{l}_M)}.$$

La ecuación (2) define el centro de masa como la media aritmética de las ubicaciones espaciales de las torres celulares en la trayectoria M.

Nuevas medidas de movilidad

Suponiendo que la región de interés se dividió en celdas de cuadrícula no superpuestas identificadas por índices en $Q \in \{1, 2, \dots, Q\}$. Denotado por l_j^{GC} la ubicación del centroide de la celda de cuadrícula $j \in Q$. Se introduce una función de mapeo $q^{GC}(\cdot)$ lo que da, para cada torre celular $i \in K$, la celda de cuadrícula $q^{GC}(i) \in Q$ de la torre a la que pertenece. Los sitios son aquellas celdas de cuadrícula que contienen al menos una torre:

$$S = \{j : j \in Q \text{ tal que exista } i \in K \text{ con } q^{GC}(i) = j\}$$

Dado que asumimos que todas las torres indexadas por K son activas en el periodo de tiempo de referencia T, S representa el conjunto de sitios en T.

Se transforma la trayectoria espaciotemporal M de (1) en la correspondiente secuencia ordenada en el tiempo de sitios a los cuales pertenecen las torres activas que aparecen en M:

$$(Ec. 3, 4.3.2) M^{GC} = \{g_1, g_2, \dots, g_n\},$$

Donde $g_i = q^{GC}(m_i) \in S$.

La medida de movilidad llamada \ "número de viajes "(NT) es un recuento del número de veces que una persona se comunica desde una celda diferente de su comunicación anterior:

$$\#\{i : i \in \{1, 2, \dots, n-1\} \text{ tal que } g_i \neq g_{i+1}\}.$$

Las otras medidas de movilidad se construyen con respecto a una red de carreteras que conecta cualquiera de los dos lugares l y l' en la región de interés. Si una localización l no está en la red de carreteras, se asume que su ubicación espacial es reemplazada por

su proyección (la ubicación a la distancia de línea recta más pequeña) en la red de carreteras. Con esta convención, asume que cualesquiera dos localizaciones l y l' están conectadas por al menos un subconjunto continuo de localizaciones llamadas la ruta de camino en la red de camino. Suponiendo que estas rutas no contienen bucles. De todas las rutas de carretera que conectan l y l' , se elige la ruta $P(l, l')$ que es más rápida, es decir, la ruta con el menor tiempo estimado de viaje. Cada ubicación en la trayectoria $P(l, l')$ debe pertenecer exactamente a una celda de cuadrícula. Denotada por $R(l, l')$ el subconjunto de Q que comprende todas las celdas de cuadrícula intersectadas por la ruta de carretera $P(l, l')$. Se definen las siguientes funciones de distancia:

(1) $d_{RD}(l, l')$ representa la longitud de $P(l, l')$

(2) $d_{TT}(l, l')$ representa el tiempo de viaje estimado entre l y l' en $P(l, l')$

(3) $d_{GC}(l, l')$ representa el número de celdas de la cuadrícula en $R(l, l')$

Se toma $d_{GC}(l, l') = d_{RD}(l, l') = d_{TT}(l, l') = 0$

El conjunto de celdas de cuadrícula visitadas asociadas con la trayectoria espaciotemporal M^{GC} desde (3) son las celdas de la cuadrícula en las rutas de carretera más rápidas que conectan sitios consecutivos en M^{GC} :

$$\mathcal{V}(M^{GC}) = \bigcup_{j=2}^n \mathcal{P}(l_{g_{j-1}}^{GC}, l_{g_j}^{GC}).$$

La medida de movilidad llamada "celdas de cuadrícula visitadas" (GCV-R) viene dada por el número de celdas de cuadrícula visitadas $\#V(M^{GC})$. La medida de movilidad llamada "sitios visitados" (SV-R) viene dada por la relación entre el número de sitios visitados y el número total de sitios en el período de tiempo de referencia T:

$$\#(\mathcal{V}(M^{GC}) \cap \mathcal{S}) / \#\mathcal{S}.$$

La medida de movilidad llamada "distancia recorrida" (DT-R) es la suma de las longitudes de las rutas más rápidas entre los sitios consecutivos desde los cuales se ha producido la comunicación:

$$\sum_{j=2}^n d_{RD} \left(l_{g_{j-1}}^{GC}, l_{g_j}^{GC} \right).$$

La medida de movilidad llamada "tiempo recorrido" (TT-R) es la suma de los tiempos de viaje estimados en las rutas más rápidas entre los sitios consecutivos de los cuales se ha producido la comunicación:

$$\sum_{j=2}^n d_{TT} \left(l_{g_{j-1}}^{GC}, l_{g_j}^{GC} \right).$$

La medida de movilidad llamada "celdas de cuadrícula viajadas" (GCT-R) es el número de celdas de cuadrícula interceptadas por las rutas de carretera más rápidas entre los sitios consecutivos de los cuales se ha producido la comunicación:

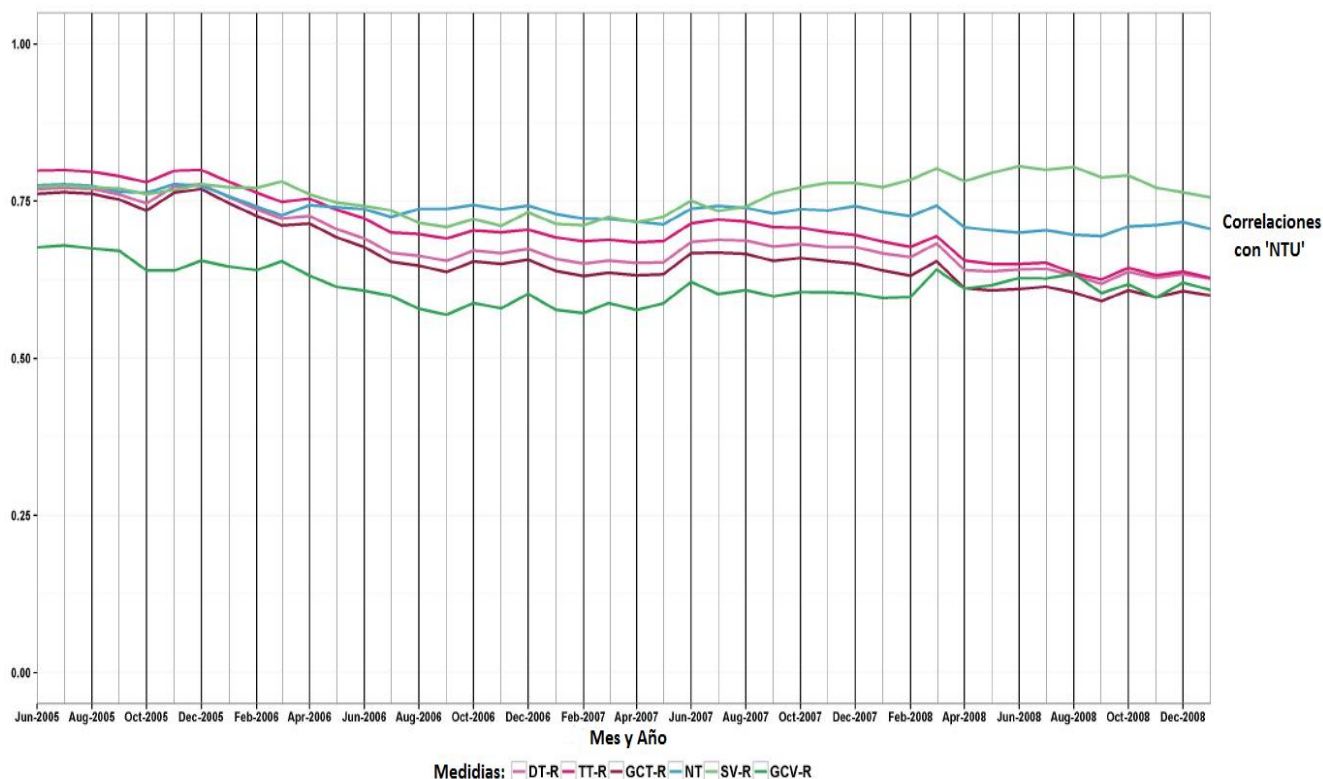
$$\sum_{j=2}^n d_{GC} \left(l_{g_{j-1}}^{GC}, l_{g_j}^{GC} \right).$$

4.3.3. RESULTADOS DE LA APLICACIÓN DE LAS MEDIDAS EN UN ANÁLISIS PREVIO [10]

Se analiza las relaciones entre las medidas existentes de movilidad (NTU, DT-SL, MDT y RoG) y las medidas propuestas (NT - grupo A, GCV-R y SV-R - grupo B, DT-R, TT-R y GCT - R - grupo C). Para los CDRs de 44 meses comprendidos entre el 1 de junio de 2005 y el 31 de enero de 2009 del país de Ruanda, se emplea las trayectorias espaciotemporales de los llamantes durante ese mes para estimar las correlaciones entre estas 10 medidas de movilidad. Se utilizó el método semiparamétrico de estimación de la cópula bayesiana gaussiana, que produce estimaciones de la matriz de correlación de la cópula gaussiana a través de un algoritmo Monte Carlo (MCMC) de cadena de Markov. Para cada uno de los 44 meses, se empleó este algoritmo MCMC para extraer 10000 muestras de la distribución posterior de la matriz de correlación de 10 dimensiones de las medidas de movilidad, después de descartar 2500 muestras como burn-in. A pesar de los grandes tamaños de muestra disponible para cada uno de los 44 meses de datos, la

estimación de parejas asociadas por coeficientes de correlación de la muestra no es ideal debido a la naturaleza no gaussiana de las distribuciones marginales univariadas de los valores de cada medida de movilidad. A pesar de los grandes tamaños de muestra disponibles para cada uno de los 44 meses de datos, la estimación de parejas asociadas por coeficientes de correlación de muestra no es ideal debido a la naturaleza no gaussiana de las distribuciones marginales univariadas de los valores de cada medida de movilidad. Estas distribuciones marginales muestran grandes proporciones de personas que llaman como de baja movilidad o que no se mueven en absoluto, y también proporciones más pequeñas pero significativas de personas que llaman tienen niveles de movilidad muy altos. Mediante el empleo de una cópula gaussiana, trata las distribuciones marginales univariadas de las medidas de movilidad como parámetros molestos, produciendo así estimaciones de correlaciones en pares con mejores propiedades estadísticas.

Figura 20. Correlaciones estimadas para cada uno de los 44 meses de las CDR entre la medida existente del número de movilidad de las torres utilizadas (NTU) y las seis medidas de los grupos A (azul), B (tonos de verde) y C (tonos de rojo).

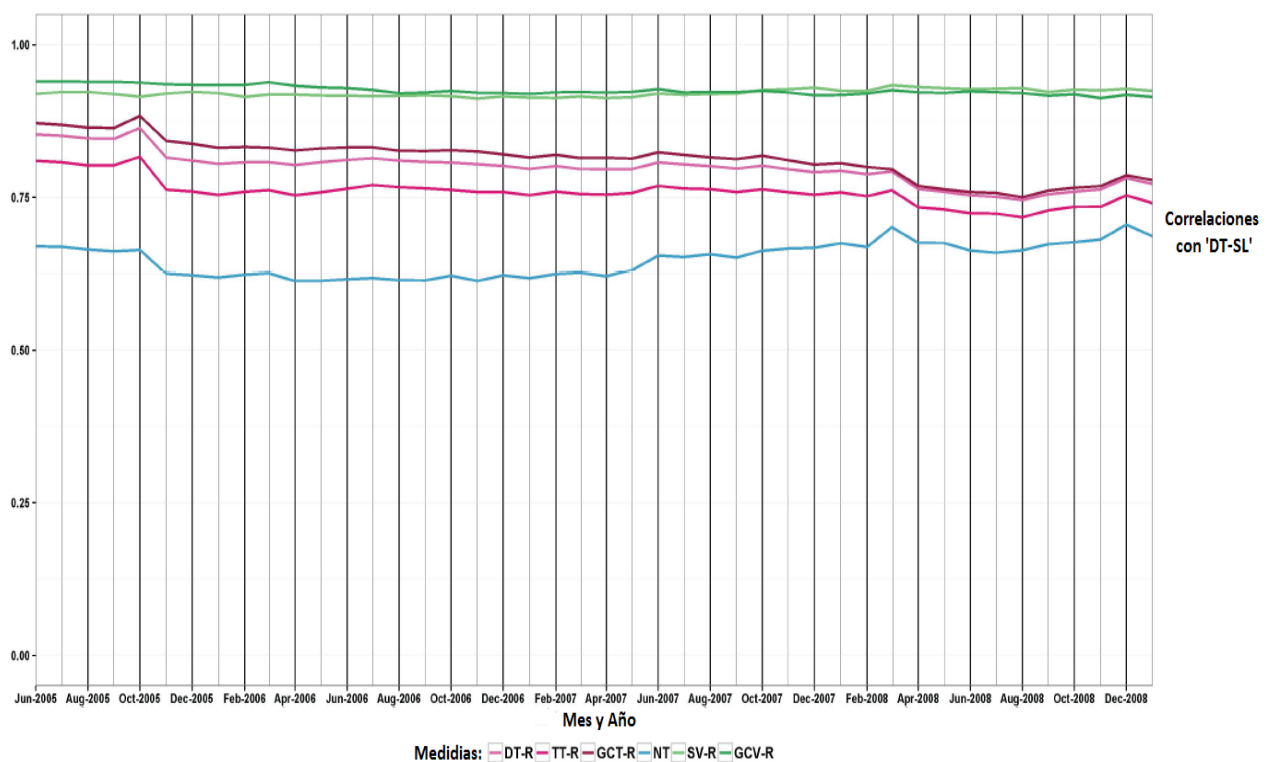


Fuente: <http://arxiv.org/abs/1408.5420>.

La figura 20 muestra que la medida existente de número de movilidad de torres utilizadas (NTU) tiene correlaciones comparables con las seis nuevas medidas. El rango de estas correlaciones parece ser mayor en 2008 comparado con 2005, posiblemente debido a que el rango espacial y la densidad de torres son significativamente más grandes en 2008. La densidad de la torre influye en NTU pero no en las nuevas medidas. Cabe destacar que NTU parece tener las asociaciones longitudinales más fuertes con la medida SV-R del grupo B que captura el rango espacial pero no captura la frecuencia de movilidad. Esto no es sorprendente, ya que NTU tampoco captura la frecuencia de la movilidad, y la forma en que captura el rango espacial se confunde por la densidad variable de las torres celulares. Sin embargo, NTU tiene la asociación más débil con la medida GCV-R, también del grupo B. Como GCV-R cuenta las celdas de la rejilla en las rutas de carretera más rápidas y SV-R sólo cuenta un subconjunto de estas celdas de

cuadrícula que también son sitios, por qué NTU sería más similar a SV-R: NTU se basa sólo en lugares en los que se hicieron llamadas. Dicho esto, las correlaciones de NTU con las seis nuevas medidas son consistentes con NTU que no se clasifican en ninguno de los grupos A, B y C.

Figura 21. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la medida existente de distancia de movilidad recorrida (DT-SL) y las seis medidas de los grupos A (azul), B (tonos de verde) y C (tonos de rojo).

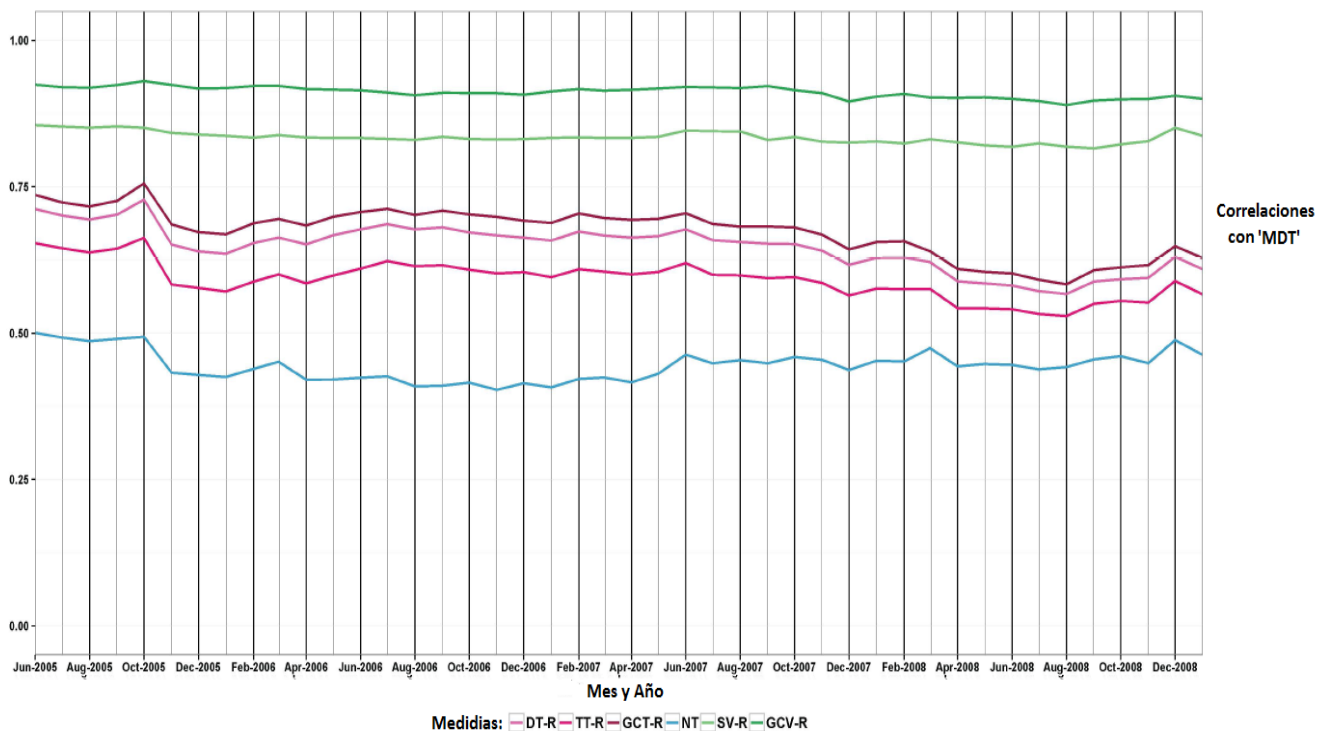


Fuente: <http://arxiv.org/abs/1408.5420>.

Las figuras 21, 22 y 23 muestran que las medidas existentes de distancia de movilidad recorrida (DT-SL), distancia máxima recorrida (MDT) y radio de giro (RoG) comparten un patrón común de asociaciones longitudinales con las seis nuevas medidas de movilidad: Sus correlaciones más fuertes son con las medidas del grupo B, sus segundas asociaciones más fuertes son con las medidas del grupo C, y sus asociaciones más débiles están con la medida del grupo A. Esto es precisamente como se esperaba en el

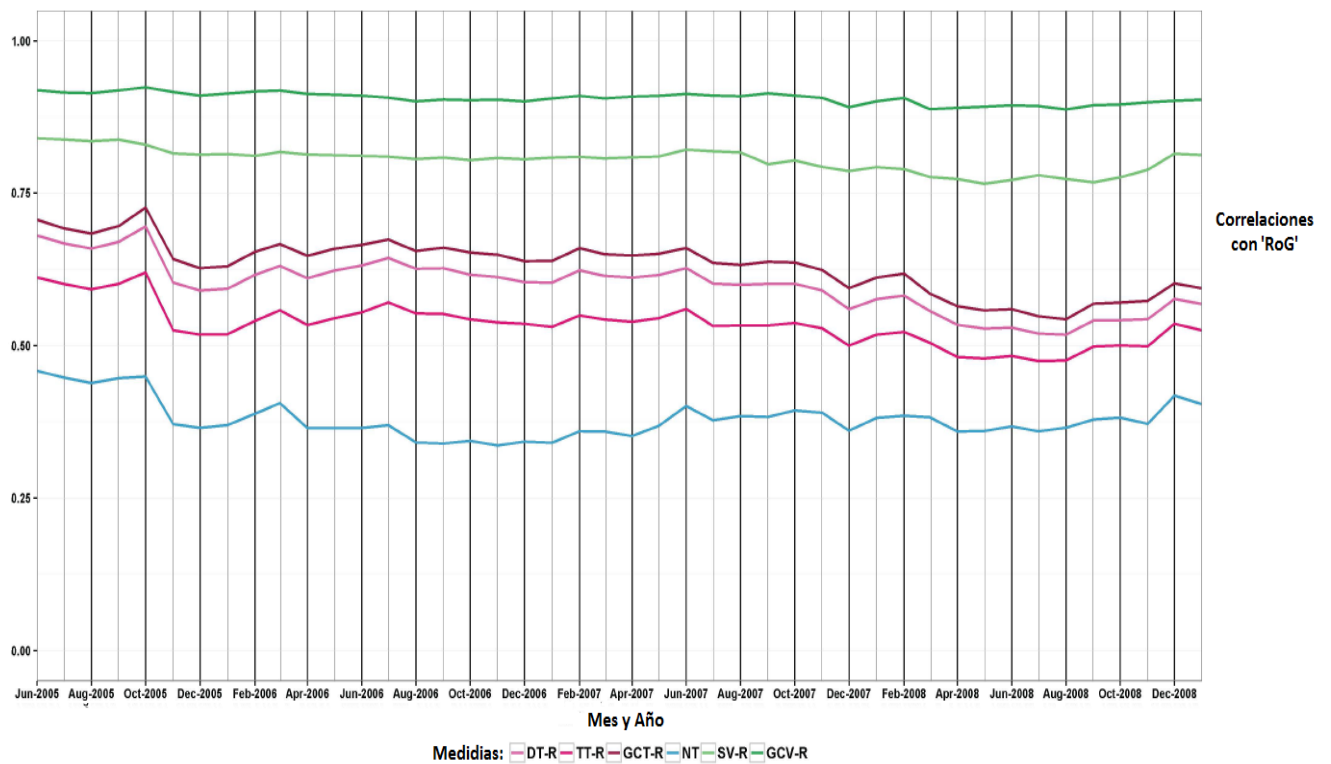
caso de MDT y RoG ya que ambas medidas capturan información relacionada con el rango espacial, pero no capturan ningún aspecto de la frecuencia de movilidad que es clave para las medidas en los grupos A y C. Sin embargo, este patrón de asociaciones es algo sorprendente para DT-SL. Clasificamos esta medida en el grupo C porque parecía captar la frecuencia de movilidad y rango espacial, y posteriormente esperábamos ver las correlaciones más grandes con las nuevas medidas de movilidad del grupo C, especialmente con la distancia recorrida (DT-R). Pero hay diferencias clave entre DT SL y DT-R: (i) DT-SL implica movimiento entre torres celulares, mientras que DT-R implica movimiento entre sitios; Y (ii) DT-SL es la suma de las distancias de la línea recta, mientras que DT-R es la suma de las distancias a través del recorrido por carretera. Las asociaciones longitudinales inferiores que observamos entre DT-SL y DT R hacen hincapié en las diferencias clave en la forma en que se construyen estas dos medidas.

Figura 22. Correlaciones estimadas para cada uno de los 44 meses de las CDR entre la medida existente de distancia máxima de movilidad recorrida (MDT) y las seis medidas de los grupos A (azul), B (tonos de verde) y C (tonos de rojo).



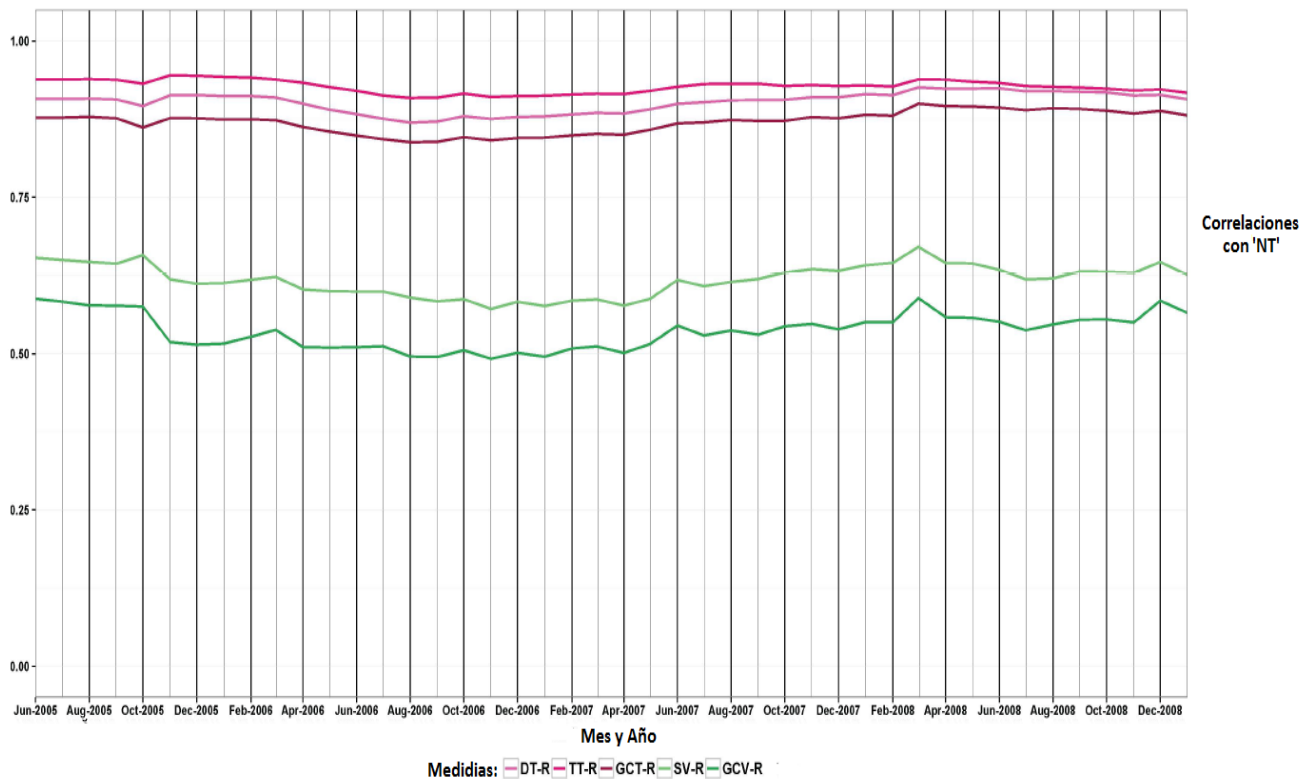
Fuente: <http://arxiv.org/abs/1408.5420>.

Figura 23. Correlaciones estimadas para cada uno de los 44 meses de CDR entre la medida existente del radio de giro de movilidad (RoG) y las seis medidas de los grupos A (azul), B (tonos de verde) y C (tonos de rojo).



Fuente: <http://arxiv.org/abs/1408.5420>.

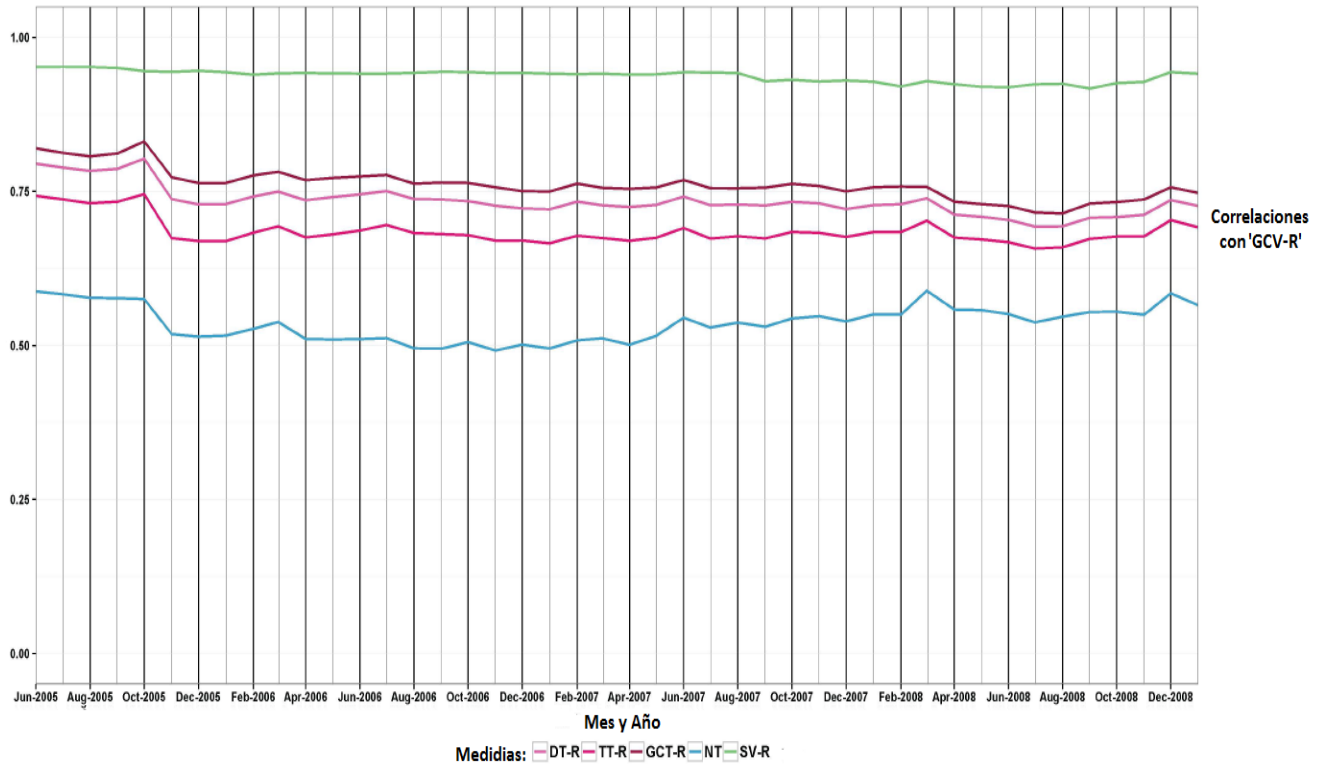
Figura 24. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la nueva medida del número de movilidad de viajes (NT) que define el grupo A y las ve medidas de los grupos B (tonos de verde) y C (tonos de rojo).



Fuente: <http://arxiv.org/abs/1408.5420>.

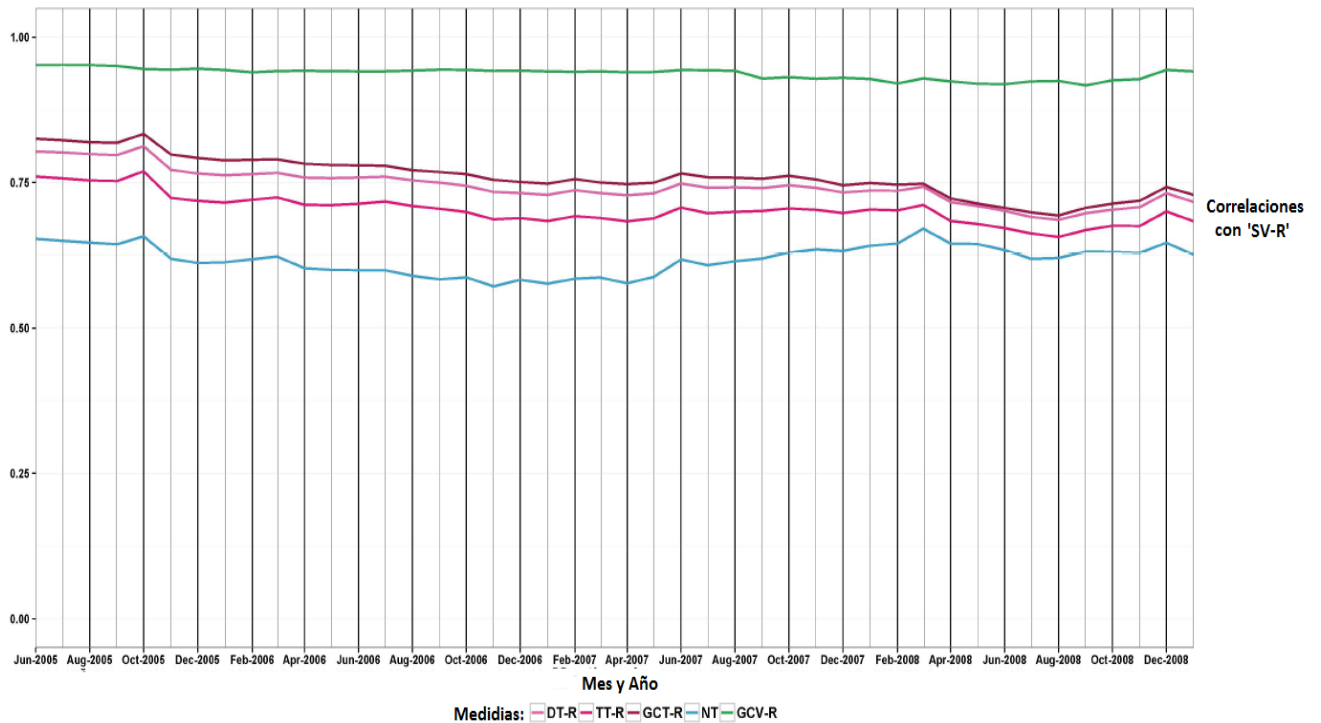
La Figura 24 muestra que la nueva medida del número de movilidad de los viajes (NT) que define el grupo A tiene las asociaciones longitudinales más fuertes con las medidas del grupo C y asociaciones más débiles con las medidas del grupo B. Esto es consistente con nuestra intuición acerca de estos El NT capta la frecuencia de movilidad, al igual que las medidas del grupo C. Pero NT no captura el rango espacial de movilidad lo que explica sus menores correlaciones con las medidas del grupo B.

Figura 25. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la nueva medida de las celdas de la cuadrícula de movilidad visitadas (GCV-R) del grupo B y la otra medida de movilidad del grupo B (SV-R, verde) Cuatro medidas de los grupos A (azul) y C (tonos de rojo).



Fuente: <http://arxiv.org/abs/1408.5420>.

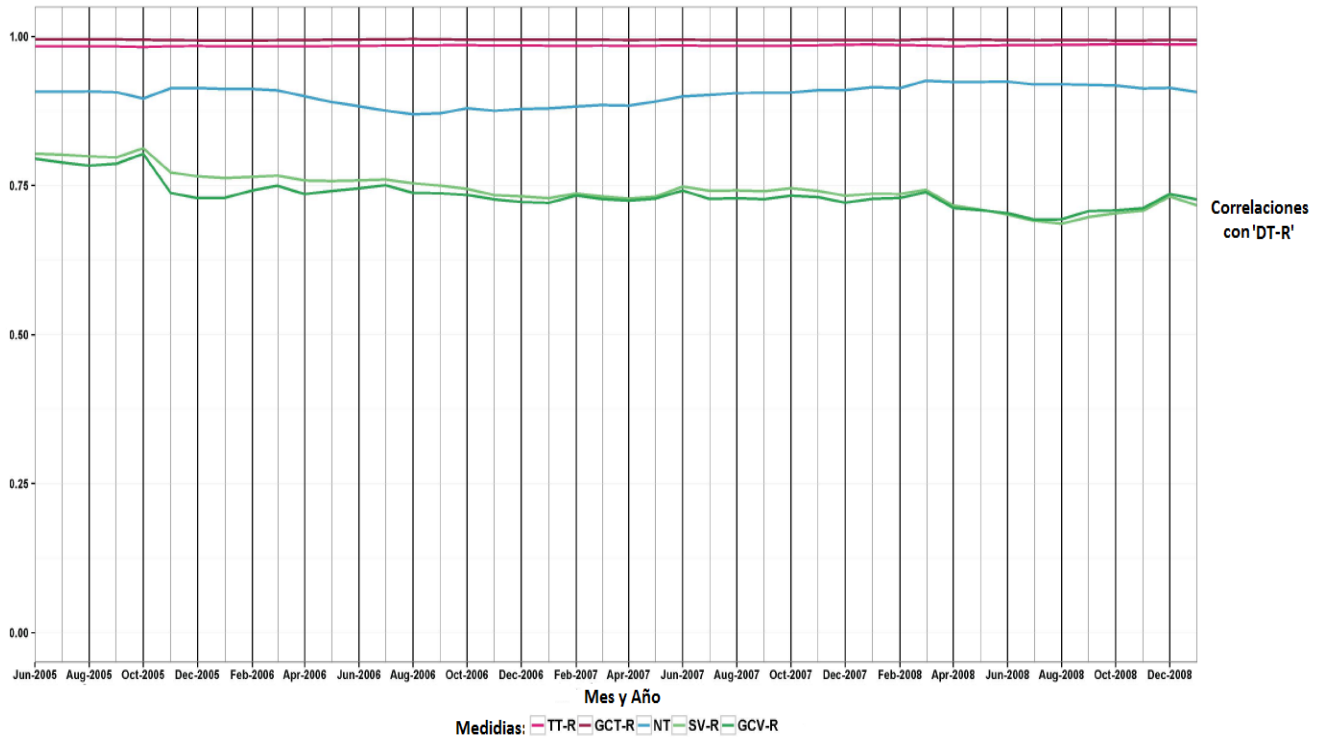
Figura 26. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la nueva medida de los sitios de movilidad visitados (SV-R) del grupo B y la otra medida de movilidad del grupo B (GCV-R, verde), así como los cuatro Medidas de los grupos A (azul) y C (tonos de rojo).



Fuente: <http://arxiv.org/abs/1408.5420>.

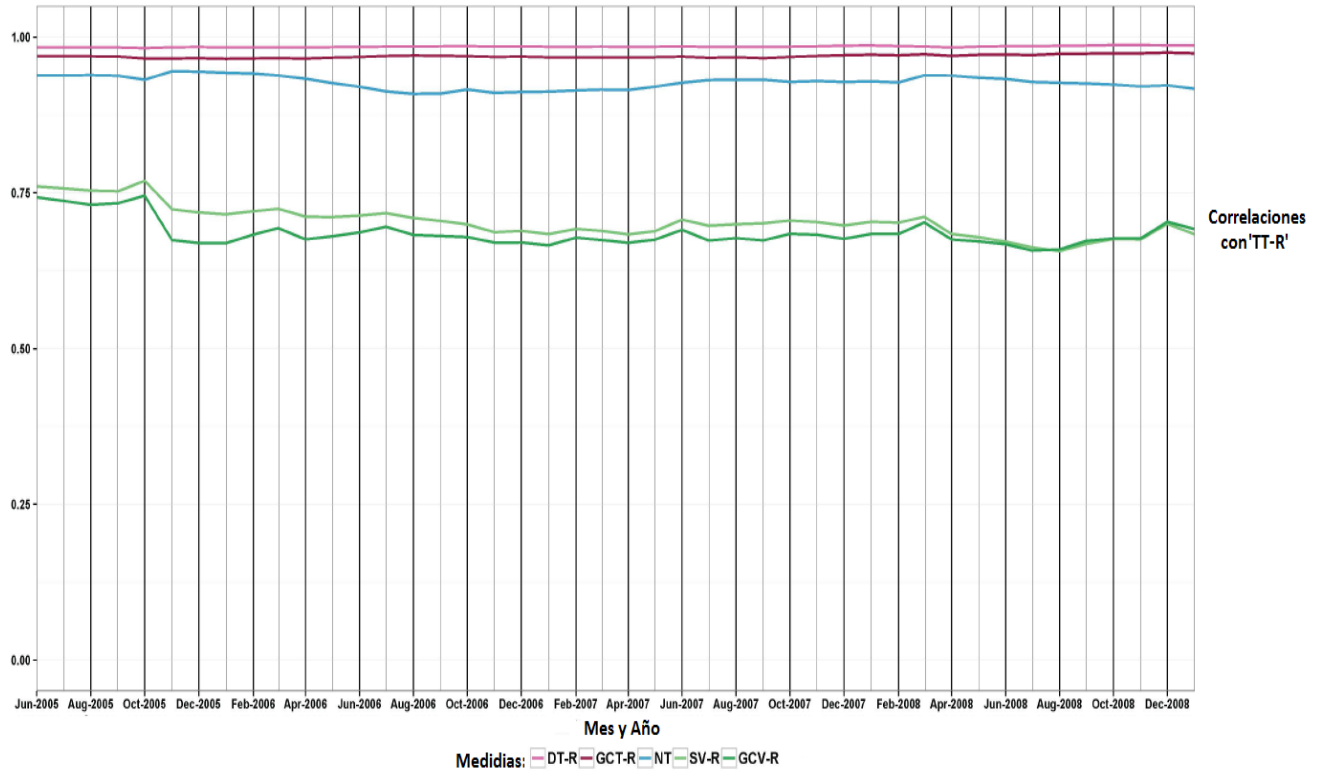
Las figuras 25 y 26 muestran que las dos nuevas medidas de movilidad del grupo B tienen fuertes correlaciones longitudinales entre sí. Sus segundas correlaciones longitudinales más fuertes se encuentran con las tres medidas del grupo C. Esto es consistente con nuestra intuición ya que las medidas del grupo B y C captan el rango espacial. Se estima que sus asociaciones longitudinales más débiles se corresponden con la medida del número de movimientos de viajes (NT) del grupo A. Esto es lógico ya que el NT captura sólo la frecuencia de movilidad, mientras que las dos medidas del grupo B sólo captan el rango espacial.

Figura 27. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la nueva medida de distancia de movilidad recorrida (DT-R) del grupo C y las otras medidas de movilidad del grupo C (TT-R y GCT-R, tonos de rojo), Así como las tres medidas de los grupos A (azul) y B (tonos de verde).



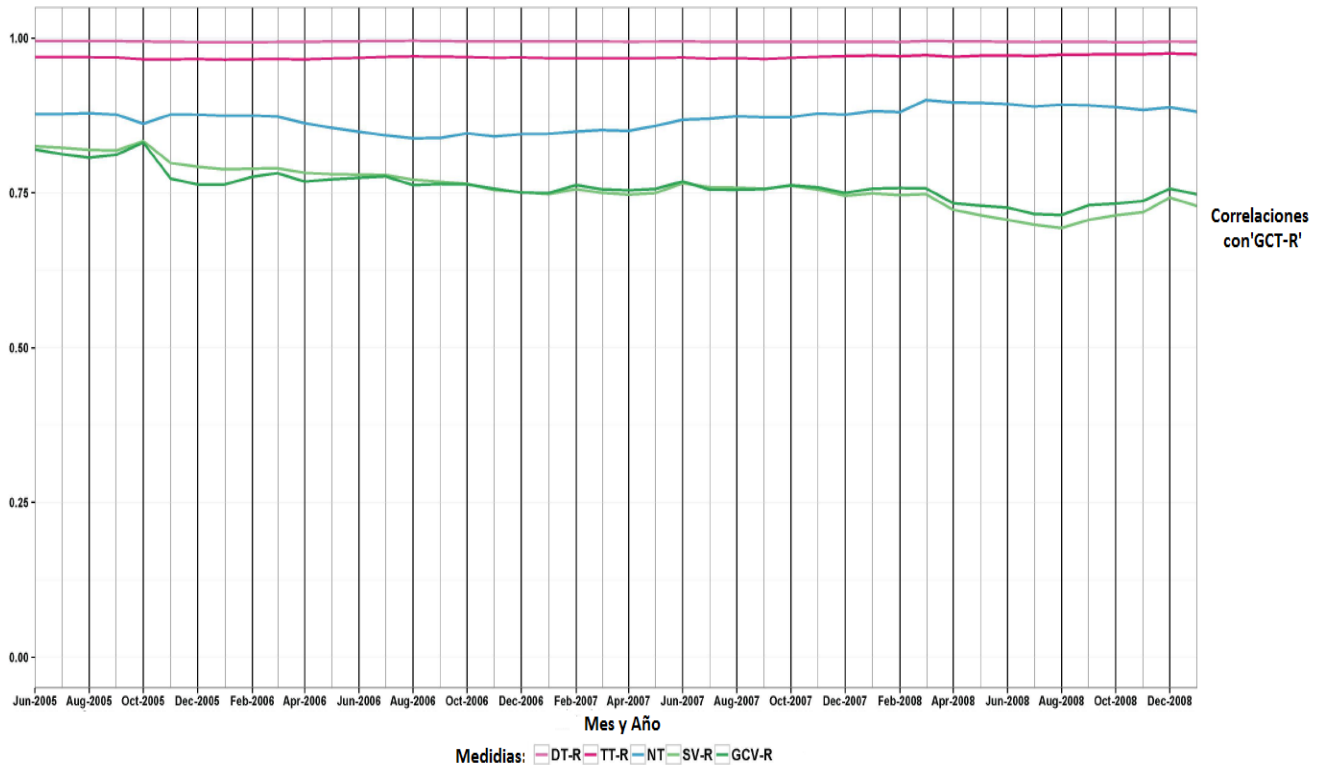
Fuente: <http://arxiv.org/abs/1408.5420>.

Figura 28. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la nueva medida del tiempo de movilidad viajado (TT-R) del grupo C y las otras medidas de movilidad del grupo C (DT-R y GCT-R, tonos de rojo), Así como las tres medidas de los grupos A (azul) y B (tonos de verde).



Fuente: <http://arxiv.org/abs/1408.5420>.

Figura 29. Las correlaciones estimadas para cada uno de los 44 meses de CDR entre la nueva medida de las celdas de la cuadrícula de movilidad viajaron (GCT-R) del grupo C y las otras medidas de movilidad del grupo C (DT-R y TT-R), así como las tres medidas de los grupos A (azul) y B (tonos de verde).



Fuente: <http://arxiv.org/abs/1408.5420>.

Las figuras 27, 28 y 29 muestran que las tres nuevas medidas de movilidad del grupo C tienen fuertes asociaciones longitudinales entre sí. Las segundas asociaciones longitudinales más fuertes están con la medida del número de movilidad de viajes (NT) del grupo A. Sus asociaciones longitudinales más débiles se estiman con las dos medidas del grupo B. Las medidas del grupo B captan el intervalo espacial como lo hacen las medidas de Grupo C, pero no captan la frecuencia de la movilidad. Aquí es interesante ver que la frecuencia de la movilidad es una de las dos dimensiones clave de la movilidad que conduce a correlaciones más grandes entre las medidas en el grupo C (que capturan ambas dimensiones) y las medidas en los grupos A y B (que capturan sólo uno de estas dimensiones).

El análisis de movilidad a través de los registros de llamadas se puede lograr siguiendo la una serie de pasos planteados en la estructura del análisis, en el que se puede realizar de dos maneras, en base a las cadenas de Márkov y en base a las variables caracterizadas.

En base a las cadenas de Márkov:

Aquí se valida cierta información para realizar el análisis de movilidad como es la posición del usuario, la velocidad del usuario, las listas de celdas vecinas, mapa de rutas, infraestructura vehicular etc. Con esta información se realizan una serie de ecuaciones las cuales permiten llegar a obtener un comportamiento de la movilidad.

En base a las variables caracterizadas:

Aquí se validan las variables caracterizadas de movilidad, donde cada una de ellas cumplen con una labor específica en el análisis, en donde unificando estas variables permite tener un comportamiento de movilidad.

De esta forma se propone realizar un análisis en el orden planteado, para realizar a futuro una validación práctica. De acuerdo a lo establecido, se validó de forma teórica que cada uno de los pasos propuestos se pueden llevar a cabo y existen investigaciones previas que los han validado.

4.4. PROCESO DE CURACIÓN DE DATOS PARA LA VISUALIZACIÓN [13]

A continuación, se menciona los procesos de curación de datos para la visualización de los patrones de movilidad por medio de la información que contiene los registros de llamadas.

Los sistemas urbanos eficientes y la gestión del transporte requieren una toma de decisiones sólida y práctica. Para ayudar en tal toma de decisiones, se utilizan herramientas de visualización que muestran la movilidad de las personas de una ciudad específica. Los métodos existentes dependen de encuestas costosas y largas, de miles de personas sobre sus hábitos de movilidad, el uso de datos de telefonía móvil celular es posible extraer patrones de movilidad sin invertir mucho en términos de tiempo y dinero.

Los datos celulares consisten en registros de la forma <userid>, <cell-tower lat-lon>, <timestamp>. Se compone de sólo el remitente o sólo el receptor, pero no ambos ID, a diferencia de CDR - registros detallados de llamadas - datos. Los proveedores de servicios de red tienen menos inhibiciones para compartir estos datos con terceros que los datos de CDR. Los datos celulares son menos precisos, más irregulares y más dispersos que los datos GPS. Pero es barato y abundantemente disponible, y penetra todos los sectores de la gente. Otra cosa buena es que los patrones de movilidad, capturados por trayectorias de ID de celda, son muy repetitivos, y por lo tanto son extraíbles. Por lo tanto, los patrones extraídos utilizando datos celulares serán buenos representantes de toda la población.

Los patrones de movilidad individuales pueden ser capturados utilizando Dynamic Bayesian Networks (DBN). En el proceso de visualización se diseña un DBN usando la ubicación, DoW (Día de la semana) y ToD (Hora del día). También se diseñaron otros DBN con subconjuntos de tres etapas.

Los patrones de movilidad colectiva pueden ser capturados agregando los resultados de las predicciones de lugar siguiente individuales. Hay que tener en cuenta que estos patrones son diferentes de la distribución estadística de las personas en diferentes lugares y épocas. Las distribuciones estadísticas típicamente no indican la ubicación previa (y el tiempo) de las personas en un lugar dado; ni puede decir el próximo destino (y el tiempo) de estas personas. La predicción agregada del siguiente lugar, además de darnos la información sobre la distribución de personas en diferentes lugares y tiempos, nos informa sobre los patrones de movilidad (de - a) de estas personas. Ayuda a entender el propósito de la visita.

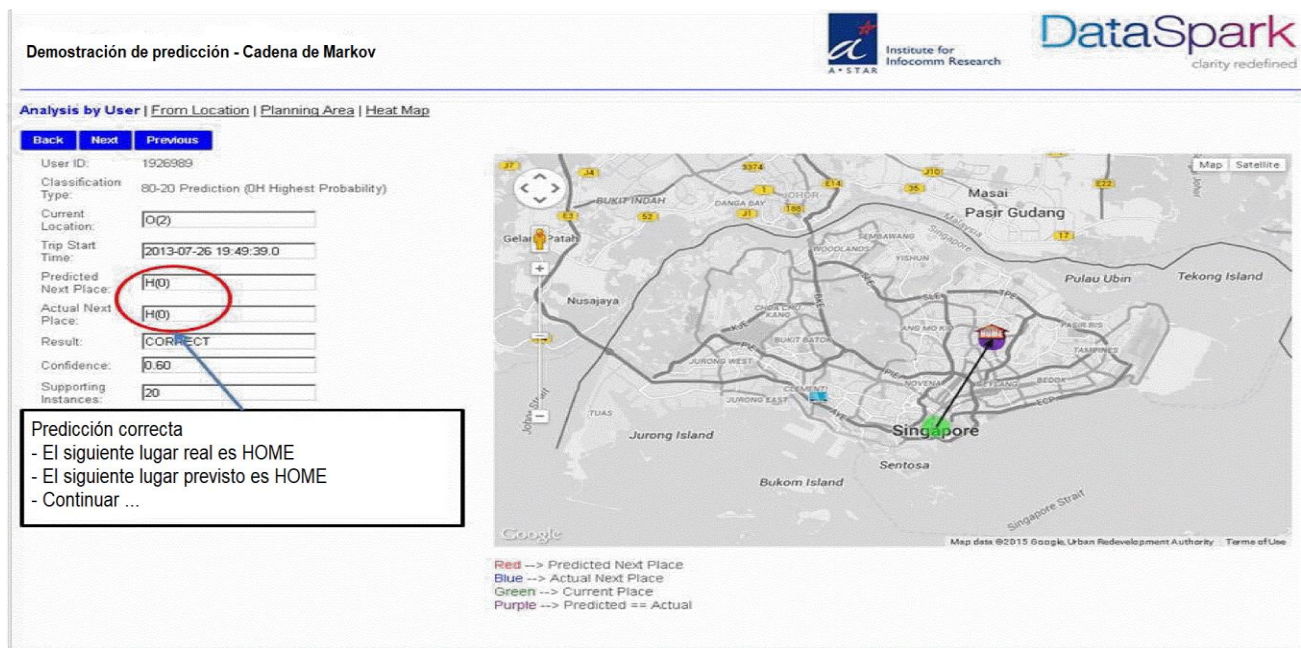
En el proceso de curación de datos, es indispensable realizar un tratamiento de los datos para llevar a cabo un proceso de visualización, que en este caso hace referencia a los patrones de movilidad en base a los registros de llamadas, en el que permita concentrarse en cualidades específicas de los datos.

El proceso se divide básicamente en tres etapas.

- Análisis de movilidad de usuarios individuales.

La figura 30 muestra una captura de pantalla para la predicción de lugar siguiente individual. En primer lugar, se extraen las regiones de estancia de un usuario. Una región de estancia es un lugar donde el usuario permanece el tiempo suficiente. Estos se extraen utilizando dos umbrales: duración de la estancia y radio. Entre éstos, el hogar (H) y el lugar de trabajo (W). Todas las otras regiones permanentes se etiquetan como 'O (.)' Donde el número dentro del paréntesis identifica la región de estadía en particular.

Figura 30. UI para la predicción de lugar siguiente individual



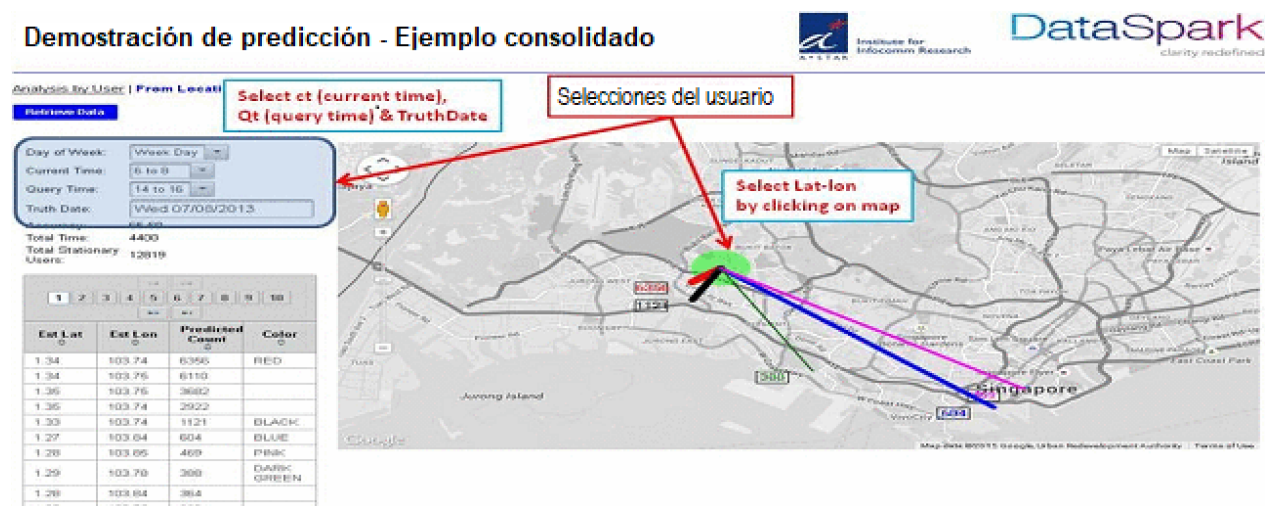
Fuente: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/7517816/>

- Análisis de movilidad agregada de usuarios.

La Figura 31 muestra una captura de pantalla de la interfaz de usuario para la predicción agregada de siguiente lugar. Implementa la consulta Q (cl, ct, qt). Para cada usuario que se encuentra en la ubicación actual (cl) y la hora actual (ct), el sistema utiliza métodos DBN para predecir el siguiente lugar en el momento de la consulta (qt). A continuación, se agrega el recuento para cada ubicación de destino (dl). Con el fin de facilitar la agregación, la ubicación actual puede ser un área de planificación o un lat-lon con un radio de 1 km, y el tiempo es una duración de dos horas. Para una agregación eficiente

sobre cl, ct, qto cualquiera de sus combinaciones, se desarrolla una estructura de índice espacio-temporal usando un índice compuesto. Un índice compuesto ofrece oportunidades para la cobertura de índices. Un índice de cobertura satisface todas las columnas solicitadas en una consulta sin realizar ninguna búsqueda adicional en el índice agrupado. Si las consultas proporcionan argumentos de búsqueda en cada una de las claves, el índice compuesto requiere menos entradas / salidas que la misma consulta utilizando un índice en cualquier atributo. Es una buena manera de hacer cumplir la unicidad de múltiples atributos. El índice compuesto consiste en ID anónimo del usuario, intervalo de tiempo, indicador de días laborables / fin de semana y cl.

Figura 31 UI para la predicción agregada del siguiente lugar



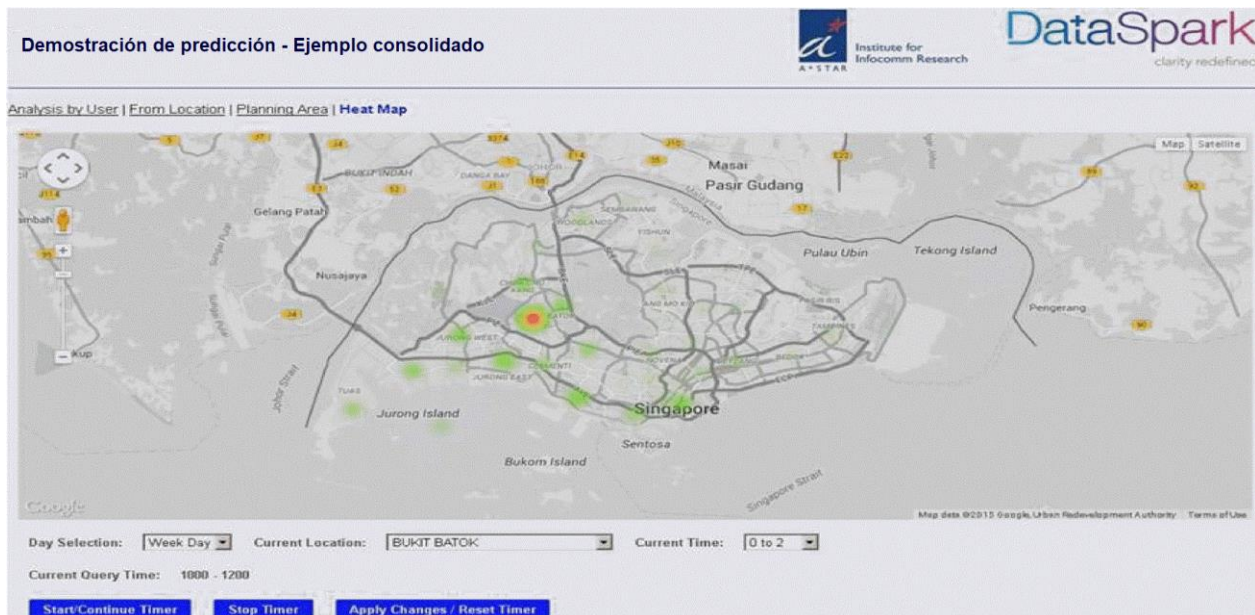
Fuente: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/7517816/>

- Analizar el mapa de calor que se muestran simultáneamente con la población en diferentes lugares y en diferentes momentos como la intensidad variable.

Los mapas de calor se trazan para mostrar la movilidad de las personas por diferentes cl y ct. La figura 32 muestra un mapa de calor donde un usuario ha seleccionado el área de planificación de Bukit Batok como cl, y 0-2 AM en un día de la semana como ct. Se trazan una serie de mapas de calor donde qt toma valores desde 2-4 AM hasta 10-12

PM. En el motor de fondo, se genera una consulta agregada Q (cl, ct, qt) para generar cada mapa de calor. La figura muestra sólo un mapa de calor tal para $qt= 1000$ a 1200 . Los colores rojo-amarillo-verde se utilizan, en orden decreciente, para representar la población variable en diferentes lugares. El círculo rojo grande es el área de planeamiento seleccionada Bukit Batok que significa que muchas personas que se encuentran en Bukit Batok en 00-02 horas también se encuentran allí en 10 -12 horas. Justo alrededor de este círculo rojo, hay una tira circular amarilla que significa un número ligeramente menor de personas. Y, finalmente, hay muchos parches verdes específicamente en las partes occidental y meridional de Singapur. De esta manera los mapas de calor ayudan a comprender la movilidad de las personas de un área de planificación durante todo un día.

Figura 32 Una instantánea de una serie de heatmaps que muestran la movilidad de las personas de bukit batok con ct fijado a la medianoche a las 2 AM mientras qt toma valores para cubrir un día entero



Fuente: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/7517816/>

La interfaz de usuario se probó a través de un conjunto de datos celular recopilado por un proveedor de servicios de red. Consta de 3,9 millones de registros anónimos de usuarios durante tres meses.

Con el fin de comparar con los métodos existentes, se han ejecutado el método a través de Nokia Mobile Data Challenge (MDC) de datos. Para 11 etiquetas, el método alcanzó el 62.79% de precisión. En el desafío del MDC, los artículos ganadores reportan menor precisión. Para 10 etiquetas, es decir, sin 'otros', el método alcanzó el 75,72% de precisión. En la literatura la mejor exactitud reportada es 73.26%. Por lo tanto, el método da la más alta precisión. Un dato interesante: la predicción para la mañana del día de la semana desde el hogar es más predecible ($Acc = 1,0$) que la predicción para la noche del día laborable desde el lugar de trabajo ($Acc = 0,22$). Uno tiene más destinos para ir en la noche que temprano en la mañana.

De esta manera se logra plantear como seria el tratamiento de los datos para llegar un proceso de visualización utilizando las redes bayesianas dinámicas (DBN).

5. CONCLUSIONES

- Las medidas de movilidad están diseñadas para ser aplicables a cualquier ámbito de investigación, desde países ricos en una infraestructura de telefonía móvil y de transporte bien desarrollada hasta los países más pobres que aún están desarrollando redes de transporte y comunicación. Constituyen un avance importante en el estudio científico social de la movilidad que podría conducir a una mejor comprensión de la salud y el bienestar humano. Estas nuevas medidas, que se derivan casi en su totalidad espacial y utilizan CDR mejoradas con datos SIG, evitan muchos de los problemas inherentes a las medidas de movilidad existentes y son independientes de la densidad de las torres celulares y de las características sociales, políticas, económicas o demográficas que influyen en la densidad de las torres. Por lo tanto, son relevantes y comparables en diferentes contextos, por ejemplo, se analiza algunas de las muchas formas en que estas nuevas medidas de movilidad basadas en el CDR pueden utilizarse para mejorar y ampliar la comprensión y la organización social.

En primer lugar, estas nuevas medidas pueden sustituir a las medidas más antiguas, a menudo basadas en encuestas por muestreo, para mejorar la comprensión de las cuestiones relacionadas con la movilidad. El beneficio es claro, dado que las medidas basadas en el CDR pueden aumentar significativamente la exactitud, detalle y el período de tiempo sobre el cual se puede registrar la movilidad. También son mucho menos costosos de obtener que las mediciones detalladas por encuesta. Los CDR pueden ser recolectados y medidas de movilidad calculadas para los encuestados que participan en las encuestas por muestreo, proporcionando al investigador no sólo un gran detalle sobre la movilidad de los encuestados, sino la oportunidad de compararla con registros de otras características y comportamientos.

En segundo lugar, estas nuevas medidas abren nuevas vías de investigación. Debido a que las CDR pueden cubrir a millones de personas, es posible calcular

medidas de movilidad a nivel de población. Por ejemplo, se puede calcular la movilidad general de una ciudad, estado, provincia o región. Esto podría compararse con los comportamientos y resultados individuales para investigar cuestiones tales como cómo la movilidad de la población influye en la migración individual. La movilidad a nivel de población también puede estar relacionada con las características de la población, como las tasas de prevalencia del VIH, las tasas de natalidad, las normas sociales, el bienestar económico o la participación política. Con las encuestas por muestreo, nunca antes fue posible calcular las características del nivel poblacional, por lo tanto, las medidas basadas en el CDR, si se calcula apropiadamente que son independientes de la densidad de las torres y las características contextuales relacionadas, crean oportunidades nuevas y posiblemente innovadoras para la movilidad.

En tercer lugar, las medidas derivadas del nivel de población derivadas de los CDRs pueden utilizarse para identificar y estudiar los casos de emergencia, como los desastres naturales y los conflictos armados. Por ejemplo, la teoría y la evidencia predicen que la gente va a cambiar sus patrones de movilidad durante y después de un terremoto o una gran explosión de una bomba. Con el acceso a datos en tiempo real, sería posible señalar un terremoto o una explosión de bomba en tiempo real, incluso en áreas remotas con malas comunicaciones y enlaces de transporte. Dado que el momento de la respuesta humanitaria influye significativamente en la magnitud y el momento del período de desastre posterior al evento, la identificación en tiempo real de los sucesos peligrosos podría, en última instancia, conducir a reducir el número de víctimas humanas de los desastres.

- Si bien las medidas basadas en el CDR pueden crear inmensas nuevas oportunidades para comprender la movilidad humana, existen varias limitaciones de las que los investigadores deben ser conscientes. Al igual que con todos los grandes datos orgánicos, la selección es una preocupación importante. Para datos

de teléfonos móviles, los usuarios de teléfonos móviles están incluidos en el conjunto de datos y los no usuarios están excluidos.

- Otra limitación clave para el uso de medidas de movilidad basadas en CDR es el error inherente. El problema principal es que, aunque las llamadas móviles se registran como ocurriendo en una torre celular, la persona que hace la llamada rara vez en esa torre. En cambio, es probable que estén a 5 o 10 km de la torre, dependiendo del tipo de antena utilizada en la torre y la topografía. Sin embargo, los beneficios de las medidas de movilidad basadas en CDR superan ampliamente a las detracciones, especialmente cuando se compara con la alternativa de medidas basadas en encuestas con error inherente debido a las dificultades humanas para recordar la localización, el tiempo y el movimiento con precisión e incapacidad para medir la movilidad a nivel poblacional.
- Mediante este tipo de análisis se puede llegar a la implementación práctica de la solución de acuerdo a la secuencia mostrada, para plantear un posible mejoramiento de la movilidad vial de una ciudad, utilizando los registros de llamada celular.
- De este análisis se derivan trabajos futuros como lo es el proyecto Eunoia: patrones de movilidad para las smart cities [11], Eunoia es un proyecto europeo de investigación del Séptimo Programa Marco, cuyo objetivo es desarrollar nuevos modelos de movilidad urbana a partir de las tecnologías de Smart Cities, los avances en Big Data y la ciencia de sistemas complejos. Uno de los temas en que se centra Eunoia es en cómo aprovechar todos los datos disponibles de redes ubicuas (registros de llamadas telefónicas, transacciones con tarjetas de crédito y huella en las redes sociales) para caracterizar patrones de movilidad en las ciudades. Las metodologías creadas se van a probar en varios pilotos en las ciudades de Londres, Barcelona y Zurich.

Otro proyecto que se deriva del análisis es, la exploración y análisis de datos de telefonía celular para estudiar comportamientos de movilidad en la ciudad de Buenos Aires [12], sobre lo que trata este proyecto es que en esta ciudad para la planificación urbana y del transporte es importante contar con herramienta que permitan comprender los patrones de movilidad de las personas en las ciudades. Un conocimiento de los flujos de movilidad que se producen en una ciudad permite tomar decisiones apropiadas de infraestructura o de políticas que brinden alternativas de transporte y que incentiven formas eficientes de movilizarse, especialmente teniendo en cuenta problemas de congestión y de contaminación ambiental que produce el transporte y la relación que tiene con la salud y la calidad de vida de la población.

BIBLIOGRAFIA

[1]. Estándar GSM (Sistema global de comunicaciones móviles), [Consulta Marzo 2017].

Disponible en:

<http://es.ccm.net/contents/681-estandar-gsm-sistema-global-de-comunicaciones-moviles>

[2] ¿Como funciona una red movil?, [Consulta Marzo 2017]. Disponible en:

<http://www.ondes-radio.orange.com/es/redes-y-estaciones-base/como-funciona-una-red-movil>.

[3]. M.C. Edith García Cárdenas, Comunicaciones móviles, [Consulta Marzo 2017].

Disponible en:

http://fcqi.tij.uabc.mx/usuarios/egarcia/CursoWEB/Comunicaciones%20Moviles/Material%20de%20Curso/Descrip_RadioBase.pdf.

[4]. Lantares solution, Big Data y la mejora de la movilidad en las grandes ciudades, España, 2003, [Consulta Octubre 2016]. Disponible en: www.lantares.com/blog/big-data-y-la-mejora-de-la-movilidad-en-las-grandes-ciudades.

[5]. José Manuel Benítez Sánchez, Big Data: Algoritmos, tecnologías y aplicaciones Universidad de Granada, España, 2016, [Consulta Octubre 2016]. Disponible en:

<http://madm.uib.es/wp-content/uploads/2016/06/Jose-Manuel-Benitez-Sanchez-Big-Data-Algoritmos-tecnologia-y-aplicaciones.pdf>.

[6] Nandish P. Kuruvatti, Hans D. Schotten, Framework to Support Mobility Context Awareness in Cellular Networks, University of Kaiserslautern, [Consulta Marzo 2017].

Disponible en: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/7504249/>.

[7] Antonio Mauttone, Héctor Cancela, María Urquhart, diseño y optimización de rutas y frecuencias en el transporte colectivo urbano, modelos y algoritmos. [Consulta Marzo 2017]. Disponible en: <https://www.fing.edu.uy/inco/pedeciba/bibliote/reptec/TR0307.pdf>

[8] Amnir Hadachi, Oleg Batrashev, Artjom Lind, Georg Singer and Eero Vainikko, Cell Phone Subscribers Mobility Prediction Using Enhanced Markov Chain Algorithm, Dearborn, Michigan, USA, [Consulta Marzo 2017]. Disponible en: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/6856442/>

[9] César Hinojosa, Algoritmo que permite obtener patrones de movilidad utilizando los datos de las tarjetas de transporte de los usuarios, Diciembre 2014, Consulta Marzo 2017]. Disponible en: <https://www.tecnocarreteras.es/2014/12/18/un-algoritmo-que-permite-obtener-patrones-de-movilidad-utilizando-los-datos-de-las-tarjetas-de-transporte-de-los-usuarios/>

[10]. Nathalie E. Williams, A. Thomas, Matthew Dunbar, Nathan Eagle, Adrian Dobra, Measures of Human Mobility Using Mobile Phone Records Enhanced with GIS Data, Cornell University Library, USA, 2014, [Consulta Marzo 2017]. Disponible en: <http://arxiv.org/abs/1408.5420>.

[11]. Sebastian Anapolsky, Exploración y análisis de datos de telefonía celular para estudiar comportamientos de movilidad en la Ciudad de Buenos Aires, Argentina, 2014, [fecha de consulta: 25 Octubre 2016]. Disponible en: https://www.academia.edu/8929162/Exploraci%C3%B3n_y_an%C3%A1lisis_de_datos_de_telefon%C3%ADa_celular_para_estudiar_comportamientos_de_movilidad_en_la_Ciudad_de_Buenos_Aires?auto=download.

[12]. Elena Tirado, Proyecto Eunoia: Patrones de movilidad para las smart cities, Universidad de Salamanca, España, 2014, [fecha de consulta: 25 Octubre 2016]. Disponible en: <http://blogthinkbig.com/eunoia-smart-cities/>.

[13]. Manoranjan Dash, Kee Kiat Koo, Shonali P. Krishnaswamy, Visualize People's Mobility - Both individually and Collectively - Using Mobile Phone Cellular Data, Institute for Infocomm Research, 2016, [fecha de consulta: Octubre de 2017]. Disponible en: <http://ieeexplore.ieee.org.ezproxy.utp.edu.co/document/7517816/>